

TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

Query Match 2.3%; Score 18; DB 1; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 8
US-08-832-877-49/c
Sequence 49, Application US/08832877
Patent No. 5840506
GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,877
FILING DATE:

CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-877-49

Query Match 2.3%; Score 18; DB 2; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 9
US-08-832-877-49/c

Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
EARLIER FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 2.3%; Score 18; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 8501 AAGGCAAAAAA 8484

RESULT 10
US-08-858-767-21
Sequence 21, Application US/08858767
Patent No. 5837468
GENERAL INFORMATION:
APPLICANT: WANG, Xun
APPLICANT: DUVICK, Jonathan P.
APPLICANT: BRIGGS, Steven P.
TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
TITLE OF INVENTION: METHOD
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,767
FILING DATE: 19-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/481,687
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/325/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

us-09-817-318-1.oli.rni

? PATENT NO.: 3853951
 ? GENERAL INFORMATION:
 ? APPLICANT: WANG, Xun
 ? APPLICANT: DUVICK, Jonathan P.
 ? APPLICANT: BRIGGS, Steven P.
 ? TITLE OF INVENTION: PCR-BASED CDNA SUBSTRUCTIVE CLONING
 ? TITLE OF INVENTION: METHOD
 ? NUMBER OF SEQUENCES: 39
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Foley & Lardner
 ? STREET: 3000 K Street, Suite 500
 ? CITY: Washington
 ? STATE: D.C.



Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	2.6	588	4	US-09-385-982-275	Sequence 275, Appl
C 2	20	2.6	176373	3	US-09-128-155-17	Sequence 17, Appl
C 3	19	2.4	6143	1	US-08-612-521-3	Sequence 3, Appl
C 4	19	2.4	7122	4	US-09-318-448-2	Sequence 2, Appl
C 5	18	2.3	776	4	US-09-372-422A-43	Sequence 43, Appl
C 6	18	2.3	3171	5	PCT-US95-09261-1	Sequence 1, Appl
C 7	18	2.3	4576	1	US-08-832-883-49	Sequence 49, Appl
C 8	18	2.3	4576	2	US-08-832-877-49	Sequence 49, Appl
C 9	18	2.3	152331	3	US-09-128-155-16	Sequence 16, Appl
C 10	17	2.2	28	2	US-08-858-767-21	Sequence 21, Appl
C 11	17	2.2	28	2	US-08-858-767-22	Sequence 22, Appl
C 12	17	2.2	28	2	US-08-858-767-23	Sequence 23, Appl
C 13	17	2.2	28	2	US-08-863-028-21	Sequence 21, Appl
C 14	17	2.2	28	2	US-08-863-028-22	Sequence 22, Appl
C 15	17	2.2	28	2	US-08-863-028-23	Sequence 23, Appl
C 16	17	2.2	418	4	US-08-642-274D-46	Sequence 46, Appl
C 17	17	2.2	418	4	US-08-952-014C-46	Sequence 46, Appl
C 18	17	2.2	545	4	US-09-328-111-506	Sequence 506, App
C 19	17	2.2	668	1	US-08-516-545-1	Sequence 1, Appl
C 20	17	2.2	854	1	US-08-684-672-25	Sequence 25, Appl
C 21	17	2.2	854	1	US-08-684-672-26	Sequence 26, Appl
C 22	17	2.2	1134	3	US-09-248-335-29	Sequence 29, Appl
C 23	17	2.2	1235	3	US-08-672-850-5	Sequence 5, Appl
C 24	17	2.2	1630	5	PCT-US92-01196-6	Sequence 6, Appl
C 25	17	2.2	1640	3	US-08-781-250-1	Sequence 1, Appl
C 26	17	2.2	1697	4	US-09-346-408-3	Sequence 5, Appl
C 27	17	2.2	1853	4	US-09-439-313-369	Sequence 369, App

EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 2.6%; Score 20; DB 3; Length 176373;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 taaggagcaaaaaa 780

Db 93715 TAAAGGCAAAAAA 93696

RESULT 3

US-08-612-521-3/C
Sequence 3, Application US/08612521
Patent No. 5786463
GENERAL INFORMATION:
APPLICANT: Peery, Robert B
APPLICANT: Skatrud, Paul L
APPLICANT: Thornevell, Susan J
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Patent Division/AEH
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,521
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hamilton, Amy E
REGISTRATION NUMBER: 33,894
REFERENCE/DOCKET NUMBER: X-9693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3169
TELEFAX: 317-276-1294
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6143 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: exon
LOCATION: 541..1072
FEATURE:
NAME/KEY: exon
LOCATION: 1130..1395
FEATURE:

NAME/KEY: exon
LOCATION: 1443..1872
FEATURE:
NAME/KEY: exon
LOCATION: 1925..2428
FEATURE:
NAME/KEY: exon
LOCATION: 2479..2863
FEATURE:
NAME/KEY: exon
LOCATION: 2913..3009
FEATURE:
NAME/KEY: exon
LOCATION: 3076..3379
FEATURE:
NAME/KEY: exon
LOCATION: 3438..3561
FEATURE:
NAME/KEY: exon
LOCATION: 3614..3776
FEATURE:
NAME/KEY: exon
LOCATION: 3831..3966
FEATURE:
NAME/KEY: exon
LOCATION: 4016..4244
FEATURE:
NAME/KEY: exon
LOCATION: 4291..4438
FEATURE:
NAME/KEY: exon
LOCATION: 4487..4743
FEATURE:
NAME/KEY: exon
LOCATION: 4793..4958
FEATURE:
NAME/KEY: exon
LOCATION: 5006..5113
FEATURE:
NAME/KEY: exon
LOCATION: 5162..5500
FEATURE:
NAME/KEY: exon
LOCATION: 5553..5588
US-08-612-521-3

Query Match 2.4%; Score 19; DB 1; Length 6143;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 aaaggcaaaaaa 780

Db 3037 AAAGCAAAAAA 3019

RESULT 4

US-09-318-448-2/C
Sequence 2, Application US/09318448
Patent No. 6210950
GENERAL INFORMATION:
APPLICANT: Johnson, William G.
APPLICANT: Stenroos, Edward S.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 7122
TYPE: DNA

; ORGANISM: Homo sapiens
US-09-318-448-2

Query Match 2.4%; Score 19; DB 4; Length 7122;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 aaaggcaaaaaa 780
|||||

Db 4104 AAAGGCAAAAAA 4086

RESULT 5

US-09-372-422A-43

; Sequence 43, Application US/09372422A

; Patent No. 6313375

; GENERAL INFORMATION:

; APPLICANT: Rudolf Jung

; APPLICANT: Francois Barriau

; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof

; FILE REFERENCE: 0919

; CURRENT APPLICATION NUMBER: US/09/372,422A

; CURRENT FILING DATE: 1999-08-11

; PRIOR APPLICATION NUMBER: US 60/098,692

; PRIOR FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 43

; LENGTH: 776

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (20)...(341)

US-09-372-422A-43

Query Match

Best Local Similarity 100.0%; Pred. No. 40; Length 776;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aaggcaaaaaa 780
|||||

Db 747 aaggcaaaaaa 764

RESULT 6

PCT-US95-09261-1

; Sequence 1, Application PC/TUS9509261

; GENERAL INFORMATION:

; APPLICANT:

; APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM

; APPLICANT: STREET: 201 West 7th Street

; APPLICANT: CITY: Austin

; APPLICANT: STATE: Texas

; APPLICANT: COUNTRY: United States of America

; APPLICANT: POSTAL CODE: 78701

; APPLICANT: TELEPHONE NO: (512)499-4462

; APPLICANT: TELEFAX: (512)499-4523

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF

; TITLE OF INVENTION: A BONE AND PROSTATE DERIVED GROWTH FACTOR

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/09261

; FILING DATE: CONCURRENTLY HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/283,701

; FILING DATE: 01-AUG-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: SERTICH, GARY J.

; REGISTRATION NUMBER: 34,430

; REFERENCE/DOCKET NUMBER: UTF422P--

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (713) 789-2679

; TELEX: 79-0924

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3171 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "DNA"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 694..2310

PCT-US95-09261-1

Query Match

Best Local Similarity 100.0%; Pred. No. 36; Score 18; DB 5; Length 3171;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aaggcaaaaaa 780
|||||

Db 3153 AAGGCAAAAAA 3170

RESULT 7

US-08-832-883-49/c

; Sequence 49, Application US/08832883

; Patent No. 5807681

; GENERAL INFORMATION:

; APPLICANT: Giordano, Antonio

; APPLICANT: Baldi, Alphonso

; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS

; TITLE OF INVENTION: OF CANCER

; NUMBER OF SEQUENCES: 115

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.

; STREET: Suite 1800 Two Penn Center Plaza

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/832,883

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Monaco, Daniel A.

; REGISTRATION NUMBER: 30,480

; REFERENCE/DOCKET NUMBER: 8321-13 US1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-8383

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:37:15 ; Search time 2133.92 Seconds
(without alignments)
7649.161 Million cell updates/sec

Title: US-09-817-318-1
Perfect score: 780
Sequence: 1 aattactgttcttcttaag.....taaggcaaaaaaaaaa 780

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	780	100.0	780	6	AX260297	Sequence
2	352	45.1	167021	9	AC079414	Homo sapi
3	34	4.4	34	6	AX260319	Sequence
4	23	2.9	23	6	AX260317	Sequence
5	23	2.9	1470	9	BC017168	Homo sapi
6	23	2.9	73590	2	AC022759	Homo sapi
7	23	2.9	180336	2	AC021731	Homo sapi
8	23	2.9	189036	9	AC006043	Homo sapi
9	22	2.8	22	6	AX260318	Sequence
10	22	2.8	37714	3	CET11F9	274042 Caenorhabdi
11	22	2.8	95008	2	AC087654	Homo sapi
12	22	2.8	110000	2	AC013622_0	AC013622 Mus muscu
13	22	2.8	110000	2	AC013623_0	AC013623 Mus muscu
14	22	2.8	113729	9	AF260011	Homo sapi
15	22	2.8	115865	9	AL358944	Human DNA
16	22	2.8	129108	9	AC011482	Homo sapi
17	22	2.8	136862	9	AC005007	Homo sapi
18	22	2.8	160566	2	AC079247	Homo sapi
19	22	2.8	181532	2	AC020700	Homo sapi
20	22	2.8	181996	2	AC068749	Homo sapi
21	22	2.8	184213	9	AC073539	Homo sapi
22	22	2.8	225009	2	AC093366	Mus muscu
23	22	2.8	252278	2	AC090126	Mus muscu
24	22	2.8	340000	9	HS21C048	AL163248 Homo sapi
25	21	2.7	733	9	HS3332860	AJ332860 Homo sapi
26	21	2.7	1234	9	AF264785	Homo sapi
27	21	2.7	1282	8	AF209988	Batrachos
28	21	2.7	1471	9	AK000415	Homo sapi
29	21	2.7	49183	9	HS0139A3	AL109916 Human DNA
30	21	2.7	55227	2	AC100331	Mus muscu
31	21	2.7	61020	6	AX251546	AX251546 Sequence
32	21	2.7	62868	2	AC016799	Homo sapi
33	21	2.7	71203	9	AL160251	Human DNA
34	21	2.7	90656	2	AC095790	Rattus no
35	21	2.7	91388	9	AC005994	Homo sapi
36	21	2.7	144223	9	AC062015	AC062015 Homo sapi
37	21	2.7	152327	2	AC026295	Homo sapi
38	21	2.7	156443	2	AC091910	Homo sapi
39	21	2.7	157257	2	AC073580	Homo sapi
40	21	2.7	162990	2	AC024707	Homo sapi
41	21	2.7	163039	2	AC010997	Homo sapi
42	21	2.7	164297	2	AC012138	Homo sapi
43	21	2.7	165412	2	AC010811	Homo sapi
44	21	2.7	169169	2	AC092527	Papio cyn
45	21	2.7	169612	2	AC027393	Homo sapi

ALIGNMENTS

RESULT 1	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	AX260297	AX260297	780 bp	DNA	linear	PAT 26-OCT-2001
DEFINITION	Sequence 1 from Patent WO0172780.					
ACCESSION	AX260297					
VERSION	AX260297.1	GI:16509264				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (sites)					
TITLE	Salceda, S., Hu, P., Recipon, H. and Cafferkey, R.					
JOURNAL	Compositions and methods of diagnosing, monitoring, staging, imaging and treating mammary gland cancer					
FEATURES	Patent: WO 0172780-A 1 04-OCT-2001; diabexus, Inc. (US)					
Source	Location/Qualifiers					
	1..780					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
BASE COUNT	267 a	115 c	119 g	279 t		
ORIGIN						

Query Match 100.0%; Score 780; DB 6; Length 780;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aattactgtctcttaagtaagccttacaccctactaaaatgtgatcaaaattttat 60
 Db 1 AATTACTGTCTCTTAAAGTAAGGCCCTACACCCTACTAAATGTGATCAAAATTTAT 60
 Qy 61 tatgaatgatgaaagcgtagctataaaattatgaagtagtattttattttatc 120
 Db 61 TATGATGATGAAAGCGTAGCTATAAAATATGAGAGTAAGTTTATTTATTTATTC 120
 Qy 121 caaatgtagtccataatagcaactcactcactaaatcttagataaaaaatgaat 180
 Db 121 CAATGTAGTTTCATAATAGCATATAGCAACTTCACCTAAATCTTAGAATAAAATGAAT 180
 Qy 181 aaaatgtaatttttggaggaatggttaatttttctacaaaattgtgacagcttt 240
 Db 181 AAAATGTTAATTTTGGAGGAATGTTAATTTTCTACAAAATTTGTGACAGCTTT 240
 Qy 241 acagaccttactctccaaattgacttgaacattacacacacaaagggtcctgtttac 300
 Db 241 ACAGACCTTACTCTTCAATGACTTGAACATTAACATCACAAGAGGTCCTGTTTAC 300
 Qy 301 aaaagaatagcaagaacttcataatgtttgtgacagctcttttcaacccttaac 360
 Db 301 AAAAGAATAGTCAAGAACTTCATGAATTTTGTACAGTGACTCTTTCTAACCTTTAATC 360
 Qy 361 caaatatttaagtgtccatctctcttcttctacaaactcatttttaactagttttc 420
 Db 361 CAATATATTTAAGTGCCATGCTCTCTCTTTATCCAACTCAATTTGTAAGTTTCT 420
 Qy 421 tctgtgagttcccttgctataaattgaagcagttctctctgaatcaccacaaactgatt 480
 Db 421 TCTGTGAGTTCCCTTGCTATAATTTGAAGCAGTTCTCTGAATCACCACCAATGATTTA 480
 Qy 481 tgaagcccatgcttttgaagatttgacactcggttgcaatctattacattgact 540
 Db 481 TGAAGCCCATGCTTTTGAAGATTGTGACTTCGGCTTTGCAATCTATTATCATTTGACT 540
 Qy 541 gtactgtcattgtctagattgtgactatcagttaggaacacacaaagatataga 600
 Db 541 GTACTTGCAATTGTTGCTAGATTGCTACTATCAGTTAGGCAATCAAAAGATATTAGA 600
 Qy 601 taatggcgaggaataaatacagaagttactgtcaataaacaagaattgtttatgggtatt 660
 Db 601 TAATGGCGAGGATAAATCAGAAGTTACTGTCAATAACAAGATTATGTTTATGGGTATT 660
 Qy 661 ttataggtgataaattcattactagcaatttcataatcatgcttttaactctcctggtt 720
 Db 661 TTATAGGTGATAAATTCATTACTAGCAATTTTCATATCATGCTTTTAAATCTCCTGGTGT 720
 Qy 721 aatatgtgactctggagactcaaatataatattgtgtaaaaggcaaaaaa 780
 Db 721 AATATGGTGACTCTGGAGACTCAATATTAATATTTGTTGTTAAAGGCCAAAAA 780

RESULT 2
 LOCUS AC079414/c 167021 bp DNA linear PRI 04-JAN-2002
 DEFINITION Homo sapiens chromosome 16 clone RP11-358L22, complete sequence.
 ACCESSION AC079414
 VERSION AC079414.5 GI:18057080
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 167021)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 167021)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 167021)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jan 4, 2002 this sequence version replaced gi:17976466.
 Sequence Quality Assessment:

This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

FEATURES

Location/Qualifiers
 1..167021
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-358L22"
 BASE COUNT 49239 a 36098 c 35465 g 46219 t
 ORIGIN

Query Match 45.1%; Score 352; DB 9; Length 167021;
 Best Local Similarity 99.4%; Pred. No. 3.3e-166;
 Matches 622; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 Qy 140 cataatgcaactcactaaatcttagaataaaaaatgaataaaatgttaatttttga 199
 Db 94271 CATAATAGCAACTTCACTAAATCTTAGAATAAAAAATGAATAAAATGTTAATTTTGA 94212
 Qy 200 ggaatggttaattttttacaaaattgtgacagctttacagaccttactctcaca 259
 Db 94211 GGAATGGTAAATTTTCTACAAAATTTGTGACAGCTTTACAGACTTACTCTTCACA 94152
 Qy 260 attgacttgaacattacacacacaaagggtcctgtttacaaaagaatagtcagaact 319
 Db 94151 ATTGACTTGAACATTAAATCAACAAAGAGGCTCTCTGTGACAAAAGAAATAGTCAAGACT 94092
 Qy 320 tcaatgaattttgacagtgactcttttcaacccttaaccataatatttaagtctc 379
 Db 94091 TCATGAATTTTGTGACAGTGACTCTTTCTAACCTTTTAATCCAAATATATTTAAGTGCTC 94032
 Qy 380 atogtctctctttacaaactcatttgttaactagtttctctctgtgagttcttgcct 439
 Db 94031 ATCGTCGTCCTTTATCAACTCATTTGTTAACTAGTTTCTTCTGTGAGTTCTCTTGCTT 93972
 Qy 440 ataattgaagcagttctctgaaatcaccccaactgattttatgaaagccatgcttttg 499
 Db 93971 ATAATTGAAGCAGTTCTCTGAAATCACCAACTGATTTTATGAAAGCCCATGCTTTTGG 93912
 Qy 500 aaagatttgcaactcggtcttgcaatctatttactgactgactgactgattgtgt 559
 Db 93911 AAAGATTGCACTTCGGCTTTCGAATCTATTACATTTGACTGTGCTTGTATTGCT 93852

KEANQAPKSGSRKPPKPRKHTRSWVLKLLLLLLLFGVAGLVACRVTELOQQPL
 CTSVNTIYDNAOGLRRHEILQWVLODSQ"

BASE COUNT
 ORIGIN

364 a 377 c 460 g 269 t

Query Match

Best Local Similarity 2.9%; Score 23; DB 9; Length 1470;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 758 gtgttaagcgcaaaaaa 780

Db 1422 GTGTAAGCGCAAAAAA 1444

RESULT 6

AC022759 73590 bp DNA linear HTG 13-JUL-2000
 AC022759 Homo sapiens clone RP11-273K8, LOW-PASS SEQUENCE SAMPLING.

AC022759.2 GI:9153403

HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 73590)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,

Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

Maconald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,F.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6922562.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5832

Center clone name: 273_K_8

* NOTE: This record contains 84 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will
 * be preserved.

1 759: contig of 759 bp in length
 760 859: gap of 100 bp
 860 1639: contig of 780 bp in length
 1640 1739: gap of 100 bp
 1740 2522: contig of 783 bp in length
 2523 2622: gap of 100 bp
 2623 3400: contig of 778 bp in length
 3401 3500: gap of 100 bp
 3501 4277: contig of 777 bp in length
 4278 4377: gap of 100 bp
 4378 5145: contig of 768 bp in length
 5146 5245: gap of 100 bp
 5246 6016: contig of 771 bp in length
 6017 6116: gap of 100 bp
 6117 6892: contig of 776 bp in length
 6893 6992: gap of 100 bp
 6993 7779: contig of 787 bp in length
 7780 7879: gap of 100 bp
 7880 8652: contig of 773 bp in length
 8653 8752: gap of 100 bp
 8753 9544: contig of 792 bp in length
 9545 9644: gap of 100 bp
 9645 10410: contig of 766 bp in length
 10411 10510: gap of 100 bp
 10511 11294: contig of 784 bp in length
 11295 11394: gap of 100 bp
 11395 12150: contig of 756 bp in length
 12151 12250: gap of 100 bp
 12251 13045: contig of 795 bp in length
 13046 13145: gap of 100 bp
 13146 13940: contig of 795 bp in length
 13941 14040: gap of 100 bp
 14041 14807: contig of 767 bp in length
 14808 14907: gap of 100 bp
 14908 15661: contig of 754 bp in length
 15662 15761: gap of 100 bp
 15762 16537: contig of 776 bp in length
 16538 16637: gap of 100 bp
 16638 17415: contig of 778 bp in length
 17416 17515: gap of 100 bp
 17516 18314: contig of 799 bp in length
 18315 18414: gap of 100 bp
 18415 19202: contig of 788 bp in length
 19203 19302: gap of 100 bp
 19303 20078: contig of 776 bp in length
 20079 20178: gap of 100 bp
 20179 20967: contig of 789 bp in length
 20968 21067: gap of 100 bp
 21068 21860: contig of 793 bp in length
 21861 21960: gap of 100 bp
 21961 22746: contig of 786 bp in length
 22747 22846: gap of 100 bp
 22847 23633: contig of 787 bp in length
 23634 23733: gap of 100 bp
 23734 24513: contig of 780 bp in length
 24514 24613: gap of 100 bp
 24614 25307: contig of 774 bp in length
 25308 25487: gap of 100 bp
 25488 26252: contig of 765 bp in length
 26253 26352: gap of 100 bp
 26353 27115: contig of 763 bp in length
 27116 27215: gap of 100 bp
 27216 27968: contig of 753 bp in length
 27969 28068: gap of 100 bp
 28069 28835: contig of 767 bp in length
 28836 28935: gap of 100 bp
 28936 29723: contig of 788 bp in length
 29724 29823: gap of 100 bp
 29824 30581: contig of 758 bp in length
 30582 30681: gap of 100 bp
 30682 31460: contig of 779 bp in length

31461 31560: gap of 100 bp
31561 32345: contig of 785 bp in length
32346 32445: gap of 100 bp
32446 32340: contig of 795 bp in length
32441 33340: gap of 100 bp
33341 34124: contig of 784 bp in length
34125 34224: gap of 100 bp
34225 35003: contig of 779 bp in length
35004 35103: gap of 100 bp
35104 35873: contig of 770 bp in length
35874 35973: gap of 100 bp
35974 36746: contig of 773 bp in length
36747 36846: gap of 100 bp
36847 37613: contig of 767 bp in length
37614 37713: gap of 100 bp
37714 38487: contig of 774 bp in length
38488 38587: gap of 100 bp
38588 39362: contig of 775 bp in length
39363 39462: gap of 100 bp
39463 40254: contig of 792 bp in length
40255 40354: gap of 100 bp
40355 41141: contig of 787 bp in length
41142 41241: gap of 100 bp
41242 42033: contig of 792 bp in length
42034 42133: gap of 100 bp
42134 42915: contig of 782 bp in length
42916 43015: gap of 100 bp
43016 43794: contig of 779 bp in length
43795 43894: gap of 100 bp
43895 44667: contig of 773 bp in length
44668 44767: gap of 100 bp
44768 45556: contig of 789 bp in length
45557 45656: gap of 100 bp
45657 46433: contig of 777 bp in length
46434 46533: gap of 100 bp
46534 47292: contig of 759 bp in length
47293 47392: gap of 100 bp
47393 48175: contig of 783 bp in length
48176 48275: gap of 100 bp
48276 49063: contig of 788 bp in length
49064 49163: gap of 100 bp
49164 49981: contig of 818 bp in length
49982 50081: gap of 100 bp
50082 50849: contig of 768 bp in length
50850 50949: gap of 100 bp
50950 51740: contig of 791 bp in length
51741 51840: gap of 100 bp
51841 52638: contig of 798 bp in length
52639 52738: gap of 100 bp
52739 53519: contig of 781 bp in length
53520 53619: gap of 100 bp
53620 54360: contig of 741 bp in length
54361 54460: gap of 100 bp
54461 55236: contig of 776 bp in length
55237 55336: gap of 100 bp
55337 56088: contig of 752 bp in length
56089 56188: gap of 100 bp
56189 56978: contig of 790 bp in length
56979 57078: gap of 100 bp
57079 57862: contig of 784 bp in length
57863 57962: gap of 100 bp
57963 58757: contig of 795 bp in length
58758 58857: gap of 100 bp
58858 59640: contig of 783 bp in length
59641 59740: gap of 100 bp
59741 60504: contig of 764 bp in length
60505 60604: gap of 100 bp
60605 61370: contig of 766 bp in length
61371 61470: gap of 100 bp
61471 62237: contig of 767 bp in length
62238 62337: gap of 100 bp

Query Match

2.9%; Score 23; DB 2; Length 73590;

Best Local Similarity 100.0%; Pred. No. 0.82; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Unordered pieces.
Qy 173 aaatgaataaaatgttaattttt 195
|||||
Db 30380 AAATGAATAAAATGTTAAATTTT 30402
RESULT 7
AC021731/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-693K15 map 11, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC021731
AC021731.3 GI:7249233
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 180336)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-693K15
Unpublished
2 (bases 1 to 180336)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye.W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6982105.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5996
Center clone name: 693_K.15
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 171270 bases at least Q40
Consensus quality: 176121 bases at least Q30
Consensus quality: 177872 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 179136; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 5507: contig of 5507 bp in length
 * 5607: gap of 100 bp
 * 12210: contig of 6603 bp in length
 * 12310: gap of 100 bp
 * 19212: contig of 6902 bp in length
 * 19312: gap of 100 bp
 * 19313: 26832: contig of 7520 bp in length
 * 26833: 26932: gap of 100 bp
 * 26933: 37970: contig of 11038 bp in length
 * 37971: 38070: gap of 100 bp
 * 38071: 47850: contig of 9780 bp in length
 * 47851: 47950: gap of 100 bp
 * 47951: 60814: contig of 12864 bp in length
 * 60815: 60914: gap of 100 bp
 * 75929: contig of 15015 bp in length
 * 75930: 76029: gap of 100 bp
 * 76030: 87989: contig of 11960 bp in length
 * 87990: 88089: gap of 100 bp
 * 88090: 104142: contig of 16053 bp in length
 * 104143: 104242: gap of 100 bp
 * 104243: 124453: contig of 20211 bp in length
 * 124454: 124553: gap of 100 bp
 * 124554: 146046: contig of 21493 bp in length
 * 146047: 146146: gap of 100 bp
 * 146147: 180336: contig of 34190 bp in length

FEATURES

```

source
1. .location/Qualifiers
1. .180336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="BP11-693K15"

```

misc_feature

misc feature

misc feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

COUNT 5740

2

Try. Match

TIME FROM ?

Query.Match 2.9%; Score 23; DB 2; Length 180336;
Best Local Similarity 100.0%; Pred. No. 0.7;

[illegible]

BASE COUNT	57400 a	31804 c	32636 g	57294 t	1202 others
ORIGIN					
/note- assembly_fragment					

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RP11-538D15; actual end is at 189036 of RP11-538D15.

FEATURES

FEATURES	Source	Query Match				DB 9;				Length 189036;			
		Best Local Similarity	100.0%;	Pred. No. 0.7;	Matches	23;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
	1. .189036												
	/organism="Homo sapiens"												
	/db_xref="taxon:9606"												
	/chromosome="7"												
	/map="7q11.23-q21.1"												
	/clone="RP11-538D15"												
	/clone_lib="RPCI-11"												
	193. .226												
repeat_region	/rpt_family="AT_rich"												
repeat_region	337. .360												
repeat_region	/rpt_family="AT_rich"												
repeat_region	436. .576												
repeat_region	/rpt_family="Alu"												
repeat_region	962. .1418												
repeat_region	/rpt_family="L1"												
repeat_region	1657. .1804												
repeat_region	/rpt_family="L2"												
repeat_region	1818. .1967												
repeat_region	/rpt_family="L1"												
repeat_region	2025. .2348												
repeat_region	/rpt_family="L1"												
repeat_region	3115. .3487												
repeat_region	/rpt_family="L1"												
repeat_region	3612. .4873												
repeat_region	/rpt_family="L1"												
repeat_region	4794. .5032												
repeat_region	/rpt_family="L1"												
repeat_region	5033. .6381												
repeat_region	/rpt_family="L1"												
repeat_region	6569. .8254												
repeat_region	/rpt_family="L1"												
repeat_region	8255. .8552												
repeat_region	/rpt_family="Alu"												
repeat_region	8553. .9044												
repeat_region	/rpt_family="L1"												
repeat_region	9054. .9354												
repeat_region	/rpt_family="Alu"												
repeat_region	9385. .9907												
repeat_region	/rpt_family="L1"												
repeat_region	10048. .10192												
repeat_region	/rpt_family="Other"												
repeat_region	11312. .11426												
repeat_region	/rpt_family="MIR"												
misc_feature	12295. .12758												
gene	/notes="match to EST AA757249 (NID:g2805112) ab69c04.sl"												
	complement(<12337..>102436)												
gene	/gene="WUGSC:H_NH0538D15.1"												
CDS	complement(join(<12337. .12975,100172. .100354, 102316. .>102436))												

```

/gene="WUGSC:H_NH0538D15.1"
/notes="KIAA0705; match to PID:g332724; H_NH0538D15.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD15413.1"
/db_xref="GI:4263738"
/translation="HINRTOFENPVLKRLQOHNPHTELGTKPLQAPGFRKPL
FTRASOLKGTFLTLTKKSNMGFTIIGDEPDEFLOVKSVIPDGPAAODGKMETG
DVIVVINEVCVLGHTHADVKLFOSVPIGOSVNLVLCRGYPLPTDPPANSMVPPLA
IMERPPVMVNGRNHYETLEYISRTSOSVPIIDTRPPHSLHSMPTDQGLDGTYPFPV
HDDNVSSASGATQAEMLTILTVKGAQGFGTIADSPGTGQRVKQLIDIQGCPGICEGG
LIVFINGQNVQNLSTHTVEVDILKDCPTGSETSLIHRG"
13527. .13729
/rpt_family="MIR"
repeat_region
14020. .14071
/rpt_family="AT_rich"
repeat_region
15108. .15599
/rpt_family="L1"
repeat_region
15674. .15875
/rpt_family="MIR"
repeat_region
16734. .16939
/rpt_family="MIR"
repeat_region
18194. .18243
/rpt_family="AT_rich"
repeat_region
19051. .19110
/rpt_family="L2"
repeat_region
19230. .19740
/rpt_family="L1"
repeat_region
19775. .19862
/rpt_family="L2"
repeat_region
20565. .20705
/rpt_family="MIR"
repeat_region
21080. .21147
/rpt_family="MIR"
repeat_region
22600. .22896
/rpt_family="Alu"
repeat_region
22979. .23265
/rpt_family="Alu"
repeat_region
24868. .24973
/rpt_family="MIR"
repeat_region
25191. .31788
/rpt_family="L1"
repeat_region
32622. .32643
/rpt_family="AT_rich"
repeat_region
33277. .33325
/rpt_family="L2"
repeat_region
34913. .35286
/rpt_family="Retroviral"
repeat_region
36922. .37087
/rpt_family="MIR"
repeat_region
37113. .37213
/rpt_family="MERL_type"
repeat_region
37248. .37270
/rpt_family="POLY_A"
repeat_region
37647. .38023
/rpt_family="MaLR"
repeat_region
38737. .38969
/rpt_family="MERL_type"
repeat_region
39425. .39525
/rpt_family="Alu"
repeat_region
40845. .40866
/rpt_family="AT_rich"
repeat_region
41861. .42012

```

Query Match 2.9%; Score 23; DB 9; Length 189036;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 758 gtgtaaagcgcaaaaaa 780

|||||

Db 178965 GTGTAAGCGCAAAAAA 178987


```

SVVSIIDDFKSSAVGLGILLGLLIIFVWDRALPSSHTFRNATHVINNESTKFM
QIIFVPRVGDDEEMKNAIGAESESEKVPAYKISPTNGOFKCTRM"
complement(join(12965..13249,13297..13639,13692..14035))
/gene="t1lf9.3"
complement(join(12965..13249,13297..13639,13692..14035))
/gene="t1lf9.3"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=123.8,
E-value=1e-33, N=1"
/codon_start=1
/protein_id="CAA98528.1"
/db_xref="GI:3879688"
/db_xref="SPTREMBL:Q22396"
/translation="MFRDAINYLENHTCLKFEYENAEATVRKNGKGYSLVGMHAG
EVQDLTLDYNCASFQTAHVEIMHAGIAHQARSDDLYLIVDSTNSNDGIENENLV
PDYGSYMLIYARDPSDKRIPIDPEYNTMGSLRVAFYDMVLLNKFYGCNHNPRKL
DCKNGYQNPANCEECLECDGFGQCLDQHEGVYLEAKKWDASGRNNYRKGIEIN
TMPEYTFALTPEGSTIEVRIKLSGFCFCQCDYNGVELKYKTDRTVSPLVCCDN
DNLWNRSTNNPFTIAKYGNNRTPHFEEFYRIPGNATAPEENN"
join(14572..14782,14825..15212,15259..15457)
/gene="t1lf9.5"
join(14572..14782,14825..15212,15259..15457)
/gene="t1lf9.5"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=92.5,
E-value=2.8e-24, N=1"
/codon_start=1
/protein_id="CAA98533.1"
/db_xref="GI:3879693"
/db_xref="SPTREMBL:Q22401"
/translation="MNYTIFFFMHIAVLNIFYRFSNGNKIVMRVGGSPETKRLEKRSK
ROALRMDNEPRPRTINVFDEORFEDNSRATVLRAMEKISNHTCIKPSPRDKRL
RIVSDKGOCAATGRVGGDOOYLSFPTSCYSVGSASLIHVIGFLSHORADREYKL
NLQRLNDFOTHOYKKYLDQWIVPYDYGISIMQYHDSNDNEYGPKSKYFRMGSOI
PSFDYLMINETYQSCGEQEINCKNRKPNPNCSECNPLGFGN"
join(16710..16914,16958..17355)
/gene="t1lf9.6"
join(16710..16914,16958..17355)
/gene="t1lf9.6"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=1.1,
E-value=1.2e-09, N=1"
/codon_start=1
/protein_id="CAA98530.1"
/db_xref="GI:3879690"
/db_xref="SPTREMBL:Q22398"
/translation="MKSFFILSIQECYKGDIVARIGRNVAEKIGGARHRROVLIR
GDEERRHKWFNTVHYFYEFNFDTVKESILRAMELISNHTCIKFSTPESEKSIKM
ESDSTTIACVAGIGVRENQLFSFNSDCYSAGVAVHELHISLGFIAHQSRDRDQYLE
FKKLDLNLQTYEQYKIWEYOELVYPYDVGVSMQYPNEE"
join(17911..17989,18042..18276,18320..18390,18435..18507,
18564..18677,18728..18905)
/gene="t1lf9.7"
join(17911..17989,18042..18276,18320..18390,18435..18507,
18564..18677,18728..18905)
/gene="t1lf9.7"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=1.1,
E-value=1.2e-09, N=1"
/codon_start=1
/protein_id="CAA98531.1"
/db_xref="GI:3879691"
/db_xref="SPTREMBL:Q22399"
/translation="MEDESSAKYPYRESMPRAERYSMKALQDEWRHLFTTAEGENNI
HNPKIPEDTNDFDDMPICNESESEITGDREVSFVHNEEDEDQESSQICESEP
OELPKLCRTDKHFKELPSGLINIKDLSLGLYLPOTLLTGRIDLNDREFAFIEV
EAEKSKELLRRRRRPLARQFSHALSRAMANLRYHYVDVTEKIERCPMRNSCKGH
LWQDMRRTAOCKNFKDPSUATVEDMMSK"
join(19342..19659,19708..19857)
/gene="t1lf9.8"
join(19342..19659,19708..19857)
/gene="t1lf9.8"
/notes="similar to zinc metalloprotease"
/codon_start=1
/protein_id="CAA98532.1"
/db_xref="GI:3879692"

```

```

/db_xref="SPTREMBL:Q22400"
/translation="MANTMGSAIVAFYDLYMINKYBECSCANNLCKNKHGYPNPSNCS
QCNCPYCGCADCORSRAEPGATEQATETHQWNTVILSDAGYRYLENKOKLQVDFIYOF
LWIMAFANKITQIRKVEKCLPGKICRGVETIKWEDPRLTSPRLCCETS"
20147..20290
/gene="t1lf9.15"
/notes="U4-1-RNA
similar to U4-1 snRNA"
20147..20290
/gene="t1lf9.15"
complement(22696..22833)
/gene="t1lf9.14"
complement(22696..22833)
/gene="t1lf9.14"
/codon_start=1
/protein_id="CAA98534.1"
/db_xref="GI:3879694"

Query Match      2.8%  Score 22;  DB 3;  Length 37714;
Best Local Similarity 100.0%;  Pred. No. 2.9;
Matches 22;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 166 gaataaaaaatgaataaatgt 187
|||||
Db 34725 GAATAAAAAATGAATAAAATGCT 34704

RESULT 11
AC087654
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP4-562D20 map 11, WORKING DRAFT.
SEQUENCE, 6 unordereded pieces.
AC087654
AC087654.2 GI:14971248
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 95008)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP4-562D20
Unpublished
2 (bases 1 to 95008)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camara,T., Campopiano,A., Choepel,I., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karakas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,K., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 20, 2001 this sequence version replaced gi:12229429.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```



```

* 11200 13868: contig of 2669 bp in length
* 13869 13968: gap of 100 bp
* 13969 18430: contig of 4462 bp in length
* 18431 18530: gap of 100 bp
* 18531 21696: contig of 3166 bp in length
* 21697 21796: gap of 100 bp
* 21797 25741: contig of 3945 bp in length
* 25742 25841: gap of 100 bp
* 25842 29357: contig of 3516 bp in length
* 29358 29457: gap of 100 bp
* 29458 32719: contig of 3262 bp in length
* 32720 32819: gap of 100 bp
* 32820 35995: contig of 3176 bp in length
* 35996 36095: gap of 100 bp
* 36096 41396: contig of 5301 bp in length
* 41397 41496: gap of 100 bp
* 41497 45153: contig of 3657 bp in length
* 45154 45253: gap of 100 bp
* 45254 49318: contig of 4065 bp in length
* 49319 49418: gap of 100 bp
* 49419 54714: contig of 5296 bp in length
* 54715 54814: gap of 100 bp
* 54815 58285: contig of 3471 bp in length
* 58286 58385: gap of 100 bp
* 58386 63447: contig of 5062 bp in length
* 63448 63547: gap of 100 bp
* 63548 67966: contig of 4419 bp in length
* 67967 68066: gap of 100 bp
* 68067 73138: contig of 5072 bp in length
* 73139 73238: gap of 100 bp
* 73239 77301: contig of 4063 bp in length
* 77302 77401: gap of 100 bp
* 77402 82081: contig of 4680 bp in length
* 82082 82181: gap of 100 bp
* 82182 86719: contig of 4538 bp in length
* 86720 86819: gap of 100 bp
* 86820 92113: contig of 5294 bp in length
* 92114 92213: gap of 100 bp
* 92214 96761: contig of 4548 bp in length
* 96762 96861: gap of 100 bp
* 96862 101458: contig of 4597 bp in length
* 101459 101558: gap of 100 bp
* 101559 107167: contig of 5609 bp in length
* 107168 107267: gap of 100 bp
* 107268 113324: contig of 6057 bp in length
* 113325 113424: gap of 100 bp
* 113425 119681: contig of 6257 bp in length
* 119682 119781: gap of 100 bp
* 119782 126056: contig of 6275 bp in length
* 126057 126156: gap of 100 bp
* 126157 132230: contig of 6074 bp in length
* 132231 132330: gap of 100 bp
* 132331 140500: contig of 8170 bp in length
* 140501 140600: gap of 100 bp
* 140601 149110: contig of 8510 bp in length
* 149111 149210: gap of 100 bp
* 149211 156954: contig of 7744 bp in length
* 156955 157054: gap of 100 bp
* 157055 163961: contig of 6907 bp in length
* 163962 164061: gap of 100 bp
* 164062 172377: contig of 8316 bp in length
* 172378 172477: gap of 100 bp
* 172478 181342: contig of 8865 bp in length
* 181343 181442: gap of 100 bp
* 181443 189465: contig of 7823 bp in length
* 189466 189365: gap of 100 bp
* 189366 197225: contig of 7860 bp in length
* 197226 197325: gap of 100 bp
* 197326 206982: contig of 9657 bp in length
* 206983 207082: gap of 100 bp
* 207083 215973: contig of 8891 bp in length
* 215974 216073: gap of 100 bp
* 216074 227424: contig of 11351 bp in length

```

```

* 227425 227524: gap of 100 bp
* 227525 237208: contig of 9684 bp in length
* 237209 237308: gap of 100 bp
* 237309 247157: contig of 9849 bp in length
* 247158 247257: gap of 100 bp
* 247258 261074: contig of 13817 bp in length
* 261075 261174: gap of 100 bp
* 261175 274135: contig of 12961 bp in length
* 274136 274235: gap of 100 bp
* 274236 289816: contig of 15581 bp in length
* 289817 289916: gap of 100 bp
* 289917 305346: contig of 15430 bp in length
* 305347 305446: gap of 100 bp
* 305447 320002: contig of 14556 bp in length
* 320003 320102: gap of 100 bp
* 320103 339946: contig of 19844 bp in length
* 339947 340046: gap of 100 bp
* 340047 366637: contig of 26591 bp in length
* 366638 366737: gap of 100 bp
* 366738 396109: contig of 29372 bp in length
* 396110 396209: gap of 100 bp
* 396210 436313: contig of 40104 bp in length.

```

FEATURES

```

source
1. .436313
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="Rp23-232H18"
/clone_lib="RPCI-23 Female Mouse BAC"

```

```

BASE COUNT 122368 a 96120 c 95329 g 116149 t 6347 others
ORIGIN

```

```

Query Match 2.8%; Score 22; DB 2: Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 759 tgtaagggcaaaaaa 780
|||||
DB 97865 TGTAAAGGCAAAAAA 97886

```

```

RESULT 13
AC013623_0/c
WPCOMMENT

```

Sequence split into 5 fragments LOCUS AC013623 Accession AC013623

Fragment Name	Begin	End
AC013623_0	1	110000
AC013623_1	100001	210000
AC013623_2	200001	310000
AC013623_3	300001	410000
AC013623_4	400001	502225

```

LOCUS AC013623 502225 bp DNA linear HTG 28-MAY-2000
DEFINITION Mus musculus clone RP23-6518, *** SEQUENCING IN PROGRESS ***, 54
unordered pieces.

```

AC013623

VERSION AC013623.5 GI:8099784

KEYWORDS HTG: HTGS_PHASE1.

SOURCE house mouse.

ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 502225)

```

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP23-6518

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 502225)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McSwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-NOV-1999), Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 28, 2000 this sequence version replaced gi:7329382.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
----- Project Information
Contact: sequence_submissions@genome.wi.mit.edu
Center project name: L3195
Center clone name: 65_I_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 15255: contig of 15255 bp in length
* 15256 15355: gap of 100 bp
* 13356 41823: contig of 26468 bp in length
* 41824 41923: gap of 100 bp
* 41924 69700: contig of 27777 bp in length
* 69701 69800: gap of 100 bp
* 69801 88128: contig of 18328 bp in length
* 88129 88228: gap of 100 bp
* 88229 120914: contig of 32686 bp in length
* 120915 121014: gap of 100 bp
* 121015 146910: contig of 25896 bp in length
* 146911 147010: gap of 100 bp
* 147011 186035: contig of 39025 bp in length
* 186036 186135: gap of 100 bp
* 186136 220111: contig of 33976 bp in length
* 220112 220211: gap of 100 bp
* 220212 222624: contig of 2413 bp in length
* 222625 222724: gap of 100 bp
* 222725 224842: contig of 2118 bp in length
* 224843 224942: gap of 100 bp
* 224943 227053: contig of 2111 bp in length
* 227054 227153: gap of 100 bp
* 227154 229756: contig of 2603 bp in length
* 229757 229856: gap of 100 bp
* 229857 232049: contig of 2193 bp in length
* 232050 232149: gap of 100 bp
* 232150 235053: contig of 2904 bp in length
* 235054 235153: gap of 100 bp
* 235154 237754: contig of 2601 bp in length
* 237755 237854: gap of 100 bp
* 237855 240561: contig of 2707 bp in length
* 240562 240661: gap of 100 bp
* 240662 244335: contig of 3674 bp in length
* 244336 244435: gap of 100 bp
* 244436 246925: contig of 2490 bp in length
* 246926 247025: gap of 100 bp
* 247026 249480: contig of 2455 bp in length
* 249481 249580: gap of 100 bp
* 249581 252655: contig of 3075 bp in length
* 252656 252755: gap of 100 bp
* 252756 256196: contig of 3441 bp in length
* 256197 256296: gap of 100 bp

* 256297 259587: contig of 3251 bp in length
* 259588 259687: gap of 100 bp
* 259688 261942: contig of 2255 bp in length
* 261943 262042: gap of 100 bp
* 262043 265148: contig of 3106 bp in length
* 265149 265248: gap of 100 bp
* 265249 268367: contig of 3119 bp in length
* 268368 268467: gap of 100 bp
* 268468 272612: contig of 4145 bp in length
* 272613 272712: gap of 100 bp
* 272713 276750: contig of 4038 bp in length
* 276751 276850: gap of 100 bp
* 276851 280557: contig of 3707 bp in length
* 280558 280657: gap of 100 bp
* 280658 284201: contig of 3544 bp in length
* 284202 284301: gap of 100 bp
* 284302 288323: contig of 4022 bp in length
* 288324 288423: gap of 100 bp
* 288424 293056: contig of 4633 bp in length
* 293057 293156: gap of 100 bp
* 293157 298341: contig of 5185 bp in length
* 298342 298441: gap of 100 bp
* 298442 303389: contig of 4948 bp in length
* 303390 303489: gap of 100 bp
* 303490 309662: contig of 6173 bp in length
* 309663 309762: gap of 100 bp
* 309763 313152: contig of 3390 bp in length
* 313153 313252: gap of 100 bp
* 313253 318191: contig of 4939 bp in length
* 318192 318291: gap of 100 bp
* 318292 322669: contig of 4378 bp in length
* 322670 322769: gap of 100 bp
* 322770 329137: contig of 6368 bp in length
* 329138 329237: gap of 100 bp
* 329238 336394: contig of 7157 bp in length
* 336395 336494: gap of 100 bp
* 336495 341574: contig of 5080 bp in length
* 341575 341674: gap of 100 bp
* 341675 347177: contig of 5503 bp in length
* 347178 347277: gap of 100 bp
* 347278 353035: contig of 5758 bp in length
* 353036 353135: gap of 100 bp
* 353136 361314: contig of 8179 bp in length
* 361315 361414: gap of 100 bp
* 361415 370891: contig of 9477 bp in length
* 370892 370991: gap of 100 bp
* 370992 379658: contig of 8667 bp in length
* 379659 379758: gap of 100 bp
* 379759 387395: contig of 7637 bp in length
* 387396 387495: gap of 100 bp
* 387496 396045: contig of 8550 bp in length
* 396046 396145: gap of 100 bp
* 396146 410181: contig of 14036 bp in length
* 410182 410281: gap of 100 bp
* 410282 426416: contig of 16135 bp in length
* 426417 426516: gap of 100 bp
* 426517 441179: contig of 14663 bp in length
* 441180 441279: gap of 100 bp
* 441280 455205: contig of 13926 bp in length
* 455206 455305: gap of 100 bp
* 455306 468105: contig of 12800 bp in length
* 468106 468205: gap of 100 bp
* 468206 481736: contig of 13531 bp in length
* 481737 481836: gap of 100 bp
* 481837 502225: contig of 20389 bp in length.

FEATURES
source

1. 502225
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-6518"

BASE COUNT 134208 a 114407 c 114480 g 132045 t 7085 others
ORIGIN


```

Query Match      2.8%; Score 22; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred No. 2.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgttaaggcgaataaaataaaaaa 780
|||||
Db 34977 TGTAAAGGCACAAAAA 34956

RESULT 14
AF260011/c
DEFINITION Homo sapiens chromosome 21 clone PAC H08866 map 21q22.1, complete
sequence.
ACCESSION AF260011
VERSION AF260011.2 GI:13621230
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113729)
REFERENCE
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhaber,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyasu,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
20289799
JOURNAL
MEDLINE
REFERENCE
AUTHORS Taudien,S., Dagand,E., Hildmann,T., Delabar,J., Blechschmidt,K.,
Schilhaber,M., Baumgart,C., Dette,M., Menzel,U., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (20-APR-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 113729)
REFERENCE
AUTHORS Taudien,S., Dagand,E., Hildmann,T., Delabar,J., Blechschmidt,K.,
Schilhaber,M., Baumgart,C., Dette,M., Menzel,U., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (13-APR-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Apr 13, 2001 this sequence version replaced gi:7677558.
FEATURES
Source
1. 113729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="PAC H08866"
BASE COUNT 37225 a 21999 c 20722 g 33783 t
ORIGIN
|||||

Query Match      2.8%; Score 22; DB 9; Length 113729;
Best Local Similarity 100.0%; Pred No. 2.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 gttattttattatcccaa 124
|||||

```

```

Db 95438 GTTTATTTATTTATCCAA 95417

RESULT 15
AL358944/c
LOCUS
DEFINITION Human DNA sequence from clone RP4-562D20 on chromosome 11 contains
ESTs, STSS and GSSs, complete sequence.
ACCESSION AL358944
VERSION AL358944.12 GI:14272299
KEYWORDS HTG; follicle stimulating hormone; FSHB.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 115865)
REFERENCE
AUTHORS Martin,S.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On May 31, 2001 this sequence version replaced gi:13446452.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; Sw;
SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP4-562D20 is
from the library RPCI-4 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VBCTOR: pCYPAC2
This sequence is the entire insert of clone RP4-562D20 The true
left end of clone RP3-414F14 is at 22396 in this sequence. The true
right end of clone RP4-815A16 is at 1704 in this sequence. This
sequence was generated by the Sanger Centre from a human chromosome
11 bacterial clone contig constructed by Gavin B. Niederfuhr A,
Schumacher N, Hummerich H, Little PF, Gessler M. Genome Res 1999
Nov;9(11):1074-86.
FEATURES
Location/Qualifiers
1. 115865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP4-562D20"
/clone_l1b="RPCI-4"
45. .630
/note="MER76 repeat: matches 32. .654 of consensus"
635. .931
/note="L1M4 repeat: matches 4749. .5048 of consensus"
935. .1895
/note="L1M4 repeat: matches 2953. .3955 of consensus"
1911. .2120
/note="AluJb repeat: matches 80. .289 of consensus"
2123. .2582
/note="L1M4 repeat: matches 2507. .2959 of consensus"
2600. .2766
/note="L1M4 repeat: matches 2212. .2372 of consensus"
4209. .4570
/note="THE1B repeat: matches 1. .364 of consensus"

```

```

repeat_region 5323..5607
/Note="AluY repeat: matches 1. .284 of consensus"
repeat_region 5629..5707
/Note="L2 repeat: matches 2627. .2702 of consensus"
repeat_region 6036..6548
/Note="MLR1D repeat: matches 2. .505 of consensus"
misc_feature 6295..6512
/Note="Single clone region. Assembly confirmed by
restriction digest data."
repeat_region 6716..7101
/Note="MLR1A2 repeat: matches 1. .374 of consensus"
repeat_region 9191..9745
/Note="MLR2A repeat: matches 1. .452 of consensus"
misc_feature 9906..10692
/Note="Sequence confirmed by AC068749 sequenced by W1BR."
repeat_region 11176..11247
/Note="36 copies 2 mer tc 84% conserved"
repeat_region 11692..11777
/Note="MLR1J repeat: matches 116. .200 of consensus"
misc_feature 11714..11788
/Note="Sequence confirmed by AC068749 sequenced by W1BR."
misc_feature 11789..11876
/Note="Sequence confirmed by AC068749 sequenced by W1BR."
misc_feature 11877..11918
/Note="Sequence confirmed by AC068749 sequenced by W1BR."
repeat_region 12975..13178
/Note="MER58A repeat: matches 4. .224 of consensus"
repeat_region 13248..13525
/Note="AluX repeat: matches 1. .305 of consensus"
repeat_region 13540..14113
/Note="LNR34 repeat: matches 1. .699 of consensus"
repeat_region 15317..15612
/Note="AluSg repeat: matches 1. .286 of consensus"
repeat_region 16123..16298
/Note="MIR repeat: matches 79. .256 of consensus"
repeat_region 16561..16647
/Note="MIR repeat: matches 5. .82 of consensus"
repeat_region 16648..17011
/Note="MLT1A1 repeat: matches 1. .365 of consensus"
misc_feature 16860..17241
/Note="Single clone region. Assembly confirmed by
restriction digest data."
repeat_region 17012..17068
/Note="MIR repeat: matches 82. .141 of consensus"
mRNA complement(join(17121..18828,20337..20538))
/Note="FSHB"
/Note="match: cDNAs: Em:AK017593 Em:AF257212 Em:AB029157
Em:AF106914 Em:X15493 Em:M36804 Em:M14853 Em:AF134151
Em:M35676 Em:AF008550 Em:E00632
match: ESTs: Em:BF704061 Em:BF702099 Em:AV754057
Em:AV755146 Em:AV754501 Em:AV753373"
/product="dJ562D20.1 (follicle stimulating hormone, beta
polypeptide)"
gene /evidence=not_experimental
complement(17121..20538)
/polyA_site /gene="FSHB"
complement(17121)
/polyA_signal /gene="FSHB"
complement(17219..17224)
/polyA_signal /gene="FSHB"
repeat_region 18136..18223
/Note="L2 repeat: matches 2581. .2675 of consensus"
misc_feature 18217..18692
/Note="match: GSS: Em:AZ082576"
misc_feature complement(18393..18595)
/Note="match: SRS: Em:C29876"
CDS complement(join(18598..18828,20337..20495))
/Note="match: proteins: Tr:Q9QYB0 Tr:Q9JYK69 Sw:Q60687
Sw:P18427 Sw:O46430 Sw:P01225 Sw:P01227 Sw:P01228
Tr:Q9RTJ9 Sw:P04837 Sw:P01226"
/codon_start=1
/evidence=not_experimental
/product="dJ562D20.1 (follicle stimulating hormone, beta
polypeptide)"
/polyA_site /db_xref="GI:15723071"
/translation="MKTIQFFFLFCWKRAICNSELNITAIIEKRCRPFQISINTT
WAGCYTRDLVYKDPARKIOKTKTEKELVYTVRVPCCAHADSLTYPYVATQCHC
GKCDSDTDCTVGRGLGPSYCSFGEMKE"
20890..21053
/Note="AluDb repeat: matches 134. .297 of consensus"
repeat_region 21813..22207
/Note="L2 repeat: matches 2244. .2628 of consensus"
repeat_region 22229..23211
/Note="L2 repeat: matches 1116. .2328 of consensus"
repeat_region 23212..23521
/Note="AluX repeat: matches 1. .310 of consensus"
repeat_region 23522..23788
/Note="L2 repeat: matches 890. .116 of consensus"
repeat_region 23800..24049
/Note="L25 copies 2 mer gg 55% conserved"
repeat_region 24113..24195
/Note="L2 repeat: matches 537. .619 of consensus"
misc_feature 24262..24541
/Note="Sequence from overlapping clone dJ414F14
(AL391623). Assembly confirmed by restriction digest."
repeat_region 24628..24878
/Note="LTR8 repeat: matches 1. .256 of consensus"
misc_feature 24649..24918
/Note="Sequence from overlapping clone dJ414F14
(AL391623). Assembly confirmed by restriction digest."
repeat_region 24954..25388
/Note="LTR8 repeat: matches 172. .691 of consensus"
repeat_region 25453..25750
/Note="AluX repeat: matches 1. .292 of consensus"
misc_feature 25586..26211
/Note="Sequence from overlapping clone dJ414F14
(AL391623). Assembly confirmed by restriction digest."
repeat_region 25597..26093
/Note="match: GSS: Em:AQ440027"
repeat_region 25751..25806
/Note="MIR repeat: matches 118. .182 of consensus"
repeat_region 27093..27261
/Note="L2 repeat: matches 2383. .2545 of consensus"
repeat_region 27463..27484
/Note="L1 copies 2 mer ta 100% conserved"
repeat_region 27582..27870
/Note="AluX repeat: matches 1. .288 of consensus"
repeat_region 28185..28497
/Note="MER33 repeat: matches 4. .324 of consensus"
repeat_region 29380..29596
/Note="MIR repeat: matches 28. .252 of consensus"
repeat_region 29895..30122
/Note="MIR repeat: matches 3. .232 of consensus"
misc_feature 30148..30186
/Note="Sequence from overlapping clone dJ414F14
(AL391623). Assembly confirmed by restriction digest."
repeat_region 30420..30499
/Note="MLT1J repeat: matches 103. .183 of consensus"
repeat_region 30600..30755

```

Query Match 2.8%; Score 22; DB 9; Length 115865;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 tttagcagtgactctttctaa 350
 |||||

Db 8342 TTGTGACAGTGACTCTTTCTAA 8321

Search completed: May 13, 2002, 23:34:41
 Job time: 7046 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 22:47:41 ; Search time 317.62 seconds
(without alignments)
4216.338 Million cell updates/sec

Title: US-09-817-318-1
Perfect score: 780
Sequence: 1 aattactgttcttcttaag.....taagcgcaaaaaaaaaa 780

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	100.0	780	AA19683	Human Mammary Gland
2	34	4.4	34	AA19706	Human MSG mam021 c
3	23	2.9	23	AA19703	Human MSG mam021 c
4	22	2.8	22	AA19704	Human MSG mam021 c
5	21	2.7	4607	AA198890	Human excretory re
6	21	2.7	4607	AA164069	Human bladder rela
7	21	2.7	61020	AA546788	Tumour suppressor
8	20	2.6	458	AAH69080	Human cervical can
9	20	2.6	490	AAH71943	Human cervical can

c	10	20	2.6	490	22	AAH72697	Human cervical can
c	11	20	2.6	550	22	AAH70743	Human cervical can
c	12	20	2.6	588	21	AAA16270	Human colon cancer
c	13	20	2.6	1092	23	ABL07199	Drosophila melanog
c	14	20	2.6	1113	21	AAAS1375	D. melanogaster dN
c	15	20	2.6	1124	22	AAAS3253	DNA encoding human
c	16	20	2.6	1352	19	AAV41918	Nucleotide sequenc
c	17	20	2.6	1518	20	AAAX9049	Human secreted pro
c	18	20	2.6	1518	22	AAAS9278	Human cDNA encodin
c	19	20	2.6	1518	24	ABA90947	Human polynucleoti
c	20	20	2.6	1877	21	AAAC9705	Human secreted pro
c	21	20	2.6	3092	23	ABL07198	Drosophila melanog
c	22	20	2.6	5866	23	ABL07306	Drosophila melanog
c	23	20	2.6	7091	23	ABL07276	Drosophila melanog
c	24	20	2.6	7128	23	ABL07308	Drosophila melanog
c	25	20	2.6	7336	23	ABL07202	Drosophila melanog
c	26	20	2.6	9021	22	AAAS46326	Tumour suppressor
c	27	20	2.6	9158	20	AAAX03047	Human IL-1ra BAC c
c	28	20	2.6	11294	23	ABL04652	Drosophila melanog
c	29	20	2.6	21890	23	ABL15986	Drosophila melanog
c	30	20	2.6	22609	22	AAAL35781	Human musculoskele
c	31	20	2.6	26928	20	AAZ32184	Human prothrombin
c	32	20	2.6	32193	22	AAAL35782	Human musculoskele
c	33	20	2.6	40862	24	ABL34072	Human immune syste
c	34	19	2.4	463	22	AAI89372	Human polynucleoti
c	35	19	2.4	481	20	AAV99731	Human adult testis
c	36	19	2.4	482	22	AAI94479	Human neuroblastom
c	37	19	2.4	498	21	AAAC94508	Cat flea hindgut a
c	38	19	2.4	606	22	AAH09208	Human cDNA clone (
c	39	19	2.4	669	10	AAAN90103	Human prealbumin c
c	40	19	2.4	789	22	AAH03248	Human cDNA clone (
c	41	19	2.4	891	22	AAI97716	Human neuroblastom
c	42	19	2.4	913	22	ABA20711	Human nervous syst
c	43	19	2.4	913	22	AAAS36162	Human cardiovascular
c	44	19	2.4	914	22	ABA20710	Human nervous syst
c	45	19	2.4	914	22	AAAS36161	Human cardiovascular

ALIGNMENTS

RESULT 1
AA19683
ID AA19683 standard; cDNA; 780 BP.
XX AC AA19683;
XX DT 18-DEC-2001 (first entry)
XX Human Mammary Gland Cancer Specific Gene (MSG) mam021 cDNA.
XX DE Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;
XX KW cancer; therapy; immune response; ss.
XX OS Homo sapiens.
XX PN WO200172780-A2.
XX PD 04-OCT-2001.
XX PF 26-MAR-2001; 2001WO-US09525.
XX PR 27-MAR-2000; 2000US-192277P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Hu P, Recipon H, Cafferkey R;
XX DR WPI; 2001-616468/71.
XX PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
PT useful for diagnosing, monitoring, staging, imaging and treating
PT mammary gland cancer

XX PS Claim 1; Page 87; 99pp; English.

CC The present sequence is human mammary gland cancer specific gene (MSG)

CC cDNA. MSG is useful for diagnosing, detecting, monitoring, staging,

CC prognosticating, imaging and treating mammary gland cancer in a patient

CC by determining the levels of MSG in cells, tissues or bodily fluids in a

CC patient and comparing the determined levels of MSG with levels of MSG

CC in cells, tissues or bodily fluids from a normal human control, where a

CC change in determined levels of MSG in the patient versus normal control

CC is associated with the presence of mammary gland cancer. MSG is used for

CC identifying potential therapeutic agents for use in imaging and treating

CC mammary gland cancer. MSG antibody conjugated to a cytotoxic agent is

CC useful for treating mammary gland cancer in a patient. MSG vaccine is

CC useful for inducing an immune response against a MSG protein and for

CC treating mammary gland cancer in a patient. MSG and its protein are

CC useful as diagnostic markers for mammary gland cancer and for diagnosis

CC and treatment of disorders of cells, tissues and organisms.

XX

SO Sequence 780 BP; 267 A; 115 C; 119 G; 279 T; 0 other;

Query Match 100.0%; Score 780; DB 22; Length 780;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aattactgttcttcttaaaagtaagccttacacccctactaaaatgtgataaaattttat 60

Db 1 aattactgttcttcttaaaagtaagccttacacccctactaaaatgtgataaaattttat 60

QY 61 tatgaatagatgaaagctgtagctataataattatgagatgaattttattttatttc 120

Db 61 tatgaatagatgaaagctgtagctataataattatgagatgaattttattttatttc 120

QY 121 caaatgtagtctcaatagcataatagcaacttcactcaaatcttagaataaaatgaat 180

Db 121 caaatgtagtctcaatagcataatagcaacttcactcaaatcttagaataaaatgaat 180

QY 181 aaatgttaatttttggaggaaatggttaattttttctacaaaattgtgacagcttt 240

Db 181 aaatgttaatttttggaggaaatggttaattttttctacaaaattgtgacagcttt 240

QY 241 acagaccttactctcaattgactgaactgaacataacatacacaagaaggctctgttac 300

Db 241 acagaccttactctcaattgactgaactgaacataacatacacaagaaggctctgttac 300

QY 301 aaagaatagtcgaagaacttcataatgttttgacagctcttttctaacacctttaac 360

Db 301 aaagaatagtcgaagaacttcataatgttttgacagctcttttctaacacctttaac 360

QY 361 caaatattatagtgctccatgctctctctctctctctctctctctctctctctctct 420

Db 361 caaatattatagtgctccatgctctctctctctctctctctctctctctctctctct 420

QY 421 tctgaggttcccttgcctataattgaagcagttctctgaatacaccacaaactgattta 480

Db 421 tctgaggttcccttgcctataattgaagcagttctctgaatacaccacaaactgattta 480

QY 481 tgaagcccatgcttttggaaaagatttgcaactcgtcttgcacattattacattgact 540

Db 481 tgaagcccatgcttttggaaaagatttgcaactcgtcttgcacattattacattgact 540

QY 541 gtacttgcatgtattgctagattgactatcatcagttagtagacaatacaaaagatttaga 600

Db 541 gtacttgcatgtattgctagattgactatcatcagttagtagacaatacaaaagatttaga 600

QY 601 taatggcagggaataatcagaagttactgtcaatacaaaagttattgttttatgggtatt 660

Db 601 taatggcagggaataatcagaagttactgtcaatacaaaagttattgttttatgggtatt 660

QY 661 ttataggtgataaaattcattactagcaaatcttcataatcatgatttttattctctggtgt 720

Db 661 ttataggtgataaaattcattactagcaaatcttcataatcatgatttttattctctggtgt 720

QY 721 aatatgtgactctggagactcaaatattataattattgtttaaaggcaaaaaaaa 780

Db 721 aatatgtgactctggagactcaaatattataattattgtttaaaggcaaaaaaaa 780

RESULT 2

AAAD19706/c

ID AAD19706 standard; DNA; 34 BP.

XX

AC AAD19706;

XX

DT 18-DEC-2001 (first entry)

XX

DE Human MSG mam021 cDNA amplifying Q-PCR probe #1.

XX

KW Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;

KW cancer; therapy; immune response; PCR probe; ss.

XX

OS Homo sapiens.

XX

PN WC200172780-A2.

XX

PD 04-OCT-2001.

XX

PF 26-MAR-2001; 2001WO-US09525.

XX

PR 27-MAR-2000; 2000US-192277P.

XX

PA (DIAD-) DIADEXUS INC.

XX

PI Salceda S, Hu P, Recipon H, Cafferkey R;

XX

WPI; 2001-616468/71.

XX

New isolated polynucleotide, mammary gland cancer specific gene (MSG), useful for diagnosing, monitoring, staging, imaging and treating mammary gland cancer

Example 3; Page 54; 99pp; English.

The present sequence is a PCR probe used for amplifying human mammary gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing, detecting, monitoring, staging, prognosticating, imaging and treating mammary gland cancer in a patient by determining the levels of MSG in cells, tissues or bodily fluids in a patient and comparing the determined levels of MSG with levels of MSG in cells, tissues or bodily fluids from a normal human control, where a change in determined levels of MSG in the patient versus normal control is associated with the presence of mammary gland cancer. MSG is used for identifying potential therapeutic agents for use in imaging and treating mammary gland cancer. MSG antibody conjugated to a cytotoxic agent is useful for treating mammary gland cancer in a patient. MSG vaccine is useful for inducing an immune response against a MSG protein and for treating mammary gland cancer in a patient. MSG and its protein are useful as diagnostic markers for mammary gland cancer and for diagnosis and treatment of disorders of cells, tissues and organisms.

Sequence 34 BP; 17 A; 6 C; 6 G; 5 T; 0 other;

Query Match

Best Local Similarity 4.4%; Score 34; DB 22; Length 34;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 gttttctctgagttccttggcctataattga 447

Db 34 GTTTTCTCTGTGAGTCTCTTTCCTATAATGA 1

RESULT 3

AAAD19703

ID AAD19703 standard; DNA; 23 BP.

```

XX AC AAD19703;
XX DT 18-DEC-2001 (first entry)
XX DE Human MSG mam021 cDNA amplifying Mam021 forward PCR primer.
XX KW Human; Mammary Gland Cancer Specific Gene: MSG; cytostatic; vaccine;
XX KW cancer; therapy; immune response; PCR primer; ss.
XX OS Homo sapiens.
XX PN WO200172780-A2.
XX PD 04-OCT-2001.
XX PF 26-MAR-2001; 2001WO-US09525.
XX PR 27-MAR-2000; 2000US-192277P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Hu P, Recipon H, Cafferkey R;
XX DR WPI; 2001-616468/71.
XX PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
XX PT useful for diagnosing, monitoring, staging, imaging and treating
XX PT mammary gland cancer -
XX PS Example 3; Page 54; 99pp; English.
XX CC The present sequence is a PCR primer used for amplifying human mammary
XX CC gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing,
XX CC detecting, monitoring, staging, prognosticating, imaging and treating
XX CC mammary gland cancer in a patient by determining the levels of MSG in
XX CC cells, tissues or bodily fluids in a patient and comparing the determined
XX CC levels of MSG with levels of MSG in cells, tissues or bodily fluids from
XX CC a normal human control, where a change in determined levels of MSG in the
XX CC patient versus normal control is associated with the presence of mammary
XX CC gland cancer. MSG is used for identifying potential therapeutic agents
XX CC for use in imaging and treating mammary gland cancer. MSG antibody
XX CC conjugated to a cytotoxic agent is useful for treating mammary gland
XX CC cancer in a patient. MSG vaccine is useful for inducing an immune
XX CC response against a MSG protein and for treating mammary gland cancer in
XX CC a patient. MSG and its protein are useful as diagnostic markers for
XX CC mammary gland cancer and for diagnosis and treatment of disorders of
XX CC cells, tissues and organisms.
XX SQ Sequence 23 BP; 4 A; 9 C; 1 G; 9 T; 0 other;

Query Match 2.9%; Score 23; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 ccacgtctcttcctttatccaact 400
Db 1 ccacgtctcttcctttatccaact 23

RESULT 4
AAD19704/c
ID AAD19704 standard; DNA; 22 BP.
XX AC AAD19704;
XX DT 18-DEC-2001 (first entry)
XX DE Human MSG mam021 cDNA amplifying Mam021 reverse PCR primer.
XX KW Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;
XX KW cancer; therapy; immune response; PCR primer; ss.

```

```

XX OS Homo sapiens.
XX PN WO200172780-A2.
XX PD 04-OCT-2001.
XX PF 26-MAR-2001; 2001WO-US09525.
XX PR 27-MAR-2000; 2000US-192277P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Hu P, Recipon H, Cafferkey R;
XX DR WPI; 2001-616468/71.
XX PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
XX PT useful for diagnosing, monitoring, staging, imaging and treating
XX PT mammary gland cancer -
XX PS Example 3; Page 54; 99pp; English.
XX CC The present sequence is a PCR primer used for amplifying human mammary
XX CC gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing,
XX CC detecting, monitoring, staging, prognosticating, imaging and treating
XX CC mammary gland cancer in a patient by determining the levels of MSG in
XX CC cells, tissues or bodily fluids in a patient and comparing the determined
XX CC levels of MSG with levels of MSG in cells, tissues or bodily fluids from
XX CC a normal human control, where a change in determined levels of MSG in the
XX CC patient versus normal control is associated with the presence of mammary
XX CC gland cancer. MSG is used for identifying potential therapeutic agents
XX CC for use in imaging and treating mammary gland cancer. MSG antibody
XX CC conjugated to a cytotoxic agent is useful for treating mammary gland
XX CC cancer in a patient. MSG vaccine is useful for inducing an immune
XX CC response against a MSG protein and for treating mammary gland cancer in
XX CC a patient. MSG and its protein are useful as diagnostic markers for
XX CC mammary gland cancer and for diagnosis and treatment of disorders of
XX CC cells, tissues and organisms.
XX SQ Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;

Query Match 2.8%; Score 22; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 agcagttctctgaaatcaccca 469
Db 22 AGCAGTTCTCTGAAATCACCCA 1

RESULT 5
AAI98890
ID AAI98890 standard; DNA; 4607 BP.
XX AC AAI98890;
XX DT 07-JAN-2002 (first entry)
XX DE Human excretory related polynucleotide SEQ ID NO 654.
XX KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
XX KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
XX KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
XX KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX KW excretory system; ds.
XX OS Homo sapiens.

```

PN WO200155313-A2.
XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US01323.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI

PR 20-OCT-2000; 2000US-02411221.
 PR 20-OCT-2000; 2000US-02411785.
 PR 20-OCT-2000; 2000US-02411786.
 PR 20-OCT-2000; 2000US-02411787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-514652/56.

Forty five bladder related polynucleotides, useful in the prevention,
 treatment and diagnosis of cancer, immune disorders, cardiovascular
 disorders and neurological diseases -

Disclosure; SEQ ID NO 102; 482pp + sequence listing; English.

XX The invention relates to forty five novel bladder related
 CC polynucleotides. The polynucleotides and the polypeptides that they
 CC encode are useful in the diagnosis, treatment and prevention of:
 CC cancer, particularly breast and ovarian cancer, and other cancers

CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal
 CC tract, liver, lung, or urogenital system; immune disorders such as
 CC Addison's disease, allergies, autoimmune haemolytic anaemia,
 CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
 CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular
 CC disorders such as myocardial ischaemias; wound healing; neurological
 CC diseases such as cerebral anoxia and epilepsy; and infectious diseases
 CC such as viral, bacterial, fungal and parasitic infections. Numerous
 CC examples of each type of disorder are given in the specification.
 CC The polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. The polynucleotides are
 CC useful for chromosome identification. They are also useful as probes
 CC for diagnosing or treating a disorder related to the female reproductive
 CC system, particularly breast and/or ovary cancer. The present
 CC invention is a bladder related polynucleotide encompassed by the
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 4607 BP; 1225 A; 1140 C; 1222 G; 1020 T; 0 other;

Query Match 2.7%; Score 21; DB 22; Length 4607;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 aaaaaatgaataaaatgttaa 190

|||||

Db 3656 aaaaaatgaataaaatgttaa 3676

RESULT 7

AAS46788

ID AAS46788 standard; DNA; 61020 BP.

XX AAS46788;

AC AAS46788;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #514.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX Homo sapiens.

OS WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2001-602752/68.

DR Fragments of chemically modified genes associated with tumour suppressor

PT genes and oncogenes, useful in designing primers and probes for

PT analysing diseases associated with cytosine methylation state e.g.

PT cancer -

XX Claim 1; SEQ ID No 514; 27pp; English.

PS

XX

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (Cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 other;

Query Match 2.7%; Score 21; DB 22; Length 61020;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 tgaataaaatgttaattttt 196
 ||||||||||||||||
 Db 58724 tgaataaaatgttaattttt 58744

RESULT 8
 AAH69080/c
 ID AAH69080 standard; cDNA; 458 BP.
 XX AC
 XX AAH69080;
 XX
 XX 19-SEP-2001 (first entry)
 XX
 XX Human cervical cancer marker nucleic acid 354.
 DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.
 XX WO200142467-A2.
 XX 14-JUN-2001.
 XX
 XX 08-DEC-2000; 2000WO-US33312.
 XX
 XX 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.
 PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Schlegel R, Deeds J, Berger A, Zhao X;
 PI WPI; 2001-375006/39.
 XX
 XX New isolated nucleic acid for diagnosing and treating cervical cancer

PT and for assessing and detecting compounds for treating the cancer -
 XX Claim 1; Page 165; 1051pp; English.
 XX
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX
 XX Sequence 458 BP; 151 A; 84 C; 76 G; 147 T; 0 other;

Query Match 2.6%; Score 20; DB 22; Length 458;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 taaagcacaataaaaaaaa 780
 ||||||||||||||||
 Db 65 TAAAGCACAATAAAAAAAA 46

RESULT 9
 AAH71943/c
 ID AAH71943 standard; cDNA; 490 BP.
 XX AC
 XX AAH71943;
 XX
 XX 19-SEP-2001 (first entry)
 XX
 XX Human cervical cancer marker nucleic acid 3217.
 DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.
 XX WO200142467-A2.
 XX 14-JUN-2001.
 XX
 XX 08-DEC-2000; 2000WO-US33312.
 XX
 XX 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.
 PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Schlegel R, Deeds J, Berger A, Zhao X;
 PI WPI; 2001-375006/39.
 XX
 XX New isolated nucleic acid for diagnosing and treating cervical cancer
 XX and for assessing and detecting compounds for treating the cancer -
 XX Claim 1; Page 633; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX
 XX Sequence 490 BP; 162 A; 94 C; 68 G; 166 T; 0 other;

```

Query Match          2.6%; Score 20; DB 22; Length 490;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
    |||||
Db 31 TAAAGGCAAAAAA 12

RESULT 10
AAH72697/C
ID AAH72697 standard; cDNA; 490 BP.
XX
AC
XX
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 3971.
XX
DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
XX
PR 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.
XX
PR 12-MAY-2000; 2000US-0203791.
XX
PR 09-JUN-2000; 2000US-0210600.
XX
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 803; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 490 BP; 162 A; 94 C; 68 G; 163 T; 3 other;

Query Match          2.6%; Score 20; DB 22; Length 490;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
    |||||
Db 31 TAAAGGCAAAAAA 12

RESULT 11
AAH70743/C
ID AAH70743 standard; cDNA; 550 BP.
XX
AC
XX
XX
DT 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #275.
XX
DE Colon cancer; detect; differential expression; human; treatment;
XX
DE detect mutation; non-invasive diagnostic method; ds.
XX
OS Homo sapiens.
XX
PN WO200012702-A2.
XX

```

```

XX
AC AAH70743;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2017.
XX
DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
XX
PR 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.
XX
PR 12-MAY-2000; 2000US-0203791.
XX
PR 09-JUN-2000; 2000US-0210600.
XX
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 430; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 550 BP; 170 A; 109 C; 91 G; 180 T; 0 other;

Query Match          2.6%; Score 20; DB 22; Length 550;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
    |||||
Db 72 TAAAGGCAAAAAA 53

RESULT 12
AAH16270/C
ID AAH16270 standard; DNA; 588 BP.
XX
AC AAH16270;
XX
XX
DT 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #275.
XX
DE Colon cancer; detect; differential expression; human; treatment;
XX
DE detect mutation; non-invasive diagnostic method; ds.
XX
OS Homo sapiens.
XX
PN WO200012702-A2.
XX

```

PD 09-MAR-2000.
 XX
 PF 30-AUG-1999; 99WO-US19424.
 XX
 PR 31-AUG-1998; 98US-0098639.
 PR 27-JAN-1999; 99US-0117393.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
 PI Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX
 DR WPI: 2000-256641/22.
 XX
 PT Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer
 XX
 PS Claim 16; Page 234; 345pp; English.
 CC This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridise to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage.
 XX
 SQ Sequence 588 BP; 187 A; 110 C; 91 G; 180 T; 20 other;

Query Match 2.6%; Score 20; DB 21; Length 588;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 761 taaggcgcaaaaaa 780
 |||||
 Db 30 TAAAGGCAAAAAA 11

RESULT 13
 ABL07199/c
 ID ABL07199 standard; cDNA: 1092 BP.
 XX
 AC ABL07199;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16079.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB63096.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX
 PS Claim 1; SEQ ID NO 16079; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1092 BP; 324 A; 259 C; 248 G; 261 T; 0 other;

Query Match 2.6%; Score 20; DB 23; Length 1092;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 323 tgaattttgacagtgcac 342
 |||||
 Db 946 TGAATTTTGCACAGTC 927

RESULT 14
 AAA51375
 ID AAA51375 standard; cDNA: 1113 BP.
 XX
 AC AAA51375;
 XX
 DT 26-SEP-2000 (first entry)
 DE D. melanogaster dNK, multisubstrate deoxyribonucleosidase cDNA.
 XX
 KW dNK; multisubstrate deoxyribonucleosidase; kinase; cytostatic;
 KW virucide; ss.
 XX
 OS Drosophila melanogaster.
 XX
 FH Key Location/Qualifiers
 FT CDS 113..865
 FT /*tag= a
 FT /product= Multisubstrate_deoxyribonucleosidase
 XX
 PN WO200036099-A1.
 XX
 PD 22-JUN-2000.
 XX
 PF 10-DEC-1999; 99WO-SE02314.
 XX
 PR 11-DEC-1998; 98SE-0004298.
 XX
 PA (KARL/) KARLSSON A.
 XX
 PI Karlsson A, Johansson M;
 XX
 DR WPI: 2000-431584/37.
 DR P-PSDB; AAY96812.
 XX
 PT New nucleic acid encoding a multisubstrate deoxyribonucleosidase with
 PT homology to a Drosophila melanogaster sequence for activating
 PT anti-viral and anti-cancer drugs to treat viral infections and cancer

```

XX PS Claim 1; Page 33; 38pp; English.
XX CC This cDNA encodes Drosophila melanogaster multibase
XX CC deoxyribonucleosidase (dNK). The multibase deoxyribonucleoside
XX CC changes nucleoside analogue prodrugs into active anti-cancer and
XX CC anti-viral drugs by phosphorylation, allowing cancer and viral
XX CC infections to be treated. Drosophila melanogaster deoxyribonucleoside
XX CC kinase is (unlike human deoxyribonucleoside kinase) a multisubstrate
XX CC enzyme and it catalyzes phosphorylation of pyrimidine and purine
XX CC deoxyribonucleosides. The catalytic rates of deoxyribonucleoside and
XX CC nucleoside analogue phosphorylation are 10 to 100 fold higher than the
XX CC maximal catalytic rates reported for mammalian enzymes, making it more
XX CC useful for activating anti-viral and anti-cancer drugs.
XX SQ Sequence 1113 BP; 324 A; 287 C; 272 G; 230 T; 0 other;

Query Match      2.6%; Score 20; DB 21; Length 1113;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
Db 1087 taaaggcaaaaaa 1106

RESULT 15
AAS33253
ID AAS33253 standard; CDNA; 1124 BP.
XX AC AAS33253;
XX AC AAS33253;
XX DT 04-DEC-2001 (first entry)
XX DE DNA encoding human secreted protein, Seq ID No 212.
XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
XX KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
XX KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
XX KW cytosolic; Alzheimer's disease; Parkinson's disease; cancer;
XX KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
XX KW Gaucher's disease; neurological disease; cerebrovascular disorder;
XX KW thrombosis; wound healing; ss.
XX OS Homo sapiens.
XX PN WO200155326-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01347.
XX PR 31-JAN-2000; 2000US-0179065.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-451931/48.
XX DR P-PSDB; AAU20544.
XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing
XX PT or treating medical conditions -
XX PS Claim 1; SEQ ID No 212; 753pp; English.
XX CC The invention relates to novel isolated nucleic acid molecules (I)
XX CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
XX CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
XX CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
XX CC the prevention, treatment and diagnosis of diseases associated with
XX CC inappropriate expression of secreted proteins. (I) and complementary

```

```

CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1124 BP; 375 A; 228 C; 238 G; 283 T; 0 other;

```

```

Query Match      2.6%; Score 20; DB 22; Length 1124;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
Db 1099 taaaggcaaaaaa 1118

```

Search completed: May 14, 2002, 00:07:17
Job time: 4776 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 22:10:05 ; Search time 1946.33 Seconds

(without alignments)

5408.962 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttctcttaaa.....taaggcaaaaaaaaaaaaaa 780

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	3.1	879	12	CNS04JBE	AL293315 Tetraodon
2	24	3.1	885	12	CNS04PKX	AL288474 Tetraodon
3	23	2.9	370	9	AI991804	AI991804 wr32606.x
4	23	2.9	481	9	AA970105	AA970105 op65f11.s
5	23	2.9	482	9	AW182002	AW182002 xj69f03.x
6	22	2.8	129	9	AI340638	AI340638 tb32c02.x
7	22	2.8	172	10	BF457830	BF457830 UI-M-B21-
8	22	2.8	333	10	BF611293	BF611293 dd78c05.y
9	22	2.8	349	9	AI706726	AI706726 UI-R-AA1-
10	22	2.8	374	9	AI927339	AI927339 wn49e04.x
11	22	2.8	393	10	BG664539	BG664539 DRABFB08
12	22	2.8	395	9	AA818458	AA818458 UI-R-A0-a
13	22	2.8	410	9	AA878990	AA878990 zf83e06.s
14	22	2.8	448	9	AA818109	AA818109 UI-R-A0-a
15	22	2.8	469	12	A2913310	A2913310 RPCI-24-1
16	22	2.8	483	9	AW046246	AW046246 UI-M-BH1-
17	22	2.8	497	9	AW529757	AW529757 UI-R-BU0-

18	2.8	520	10	BF162583	BF162583 601770225
19	2.8	532	12	AO457564	AO457564 HS.5097_B
20	2.8	545	12	BH120999	BH120999 RPCI-24-2
21	2.8	546	9	AA957399	AA957399 UI-R-E1-f
22	2.8	572	12	BH095483	BH095483 RPCI-24-2
23	2.7	227	10	BF847151	BF847151 CM3-EN004
24	2.7	297	10	BF847115	BF847115 CM3-EN004
25	2.7	328	9	AW511570	AW511570 xu60a08.x
26	2.7	328	12	AO948067	AO948067 Sheared D
27	2.7	329	9	AW469992	AW469992 xr27c12.x
28	2.7	348	10	BE950018	BE950018 UI-M-CD0-
29	2.7	387	10	BM148305	BM148305 TCAAPIT23
30	2.7	397	10	BG290862	BG290862 602387011
31	2.7	430	10	BF511890	BF511890 UI-H-B14-
32	2.7	448	10	BI158830	BI158830 602922030
33	2.7	492	10	BF395183	BF395183 UI-R-CM0-
34	2.7	495	10	BM145066	BM145066 TCAAPID14
35	2.7	678	12	AZ344764	AZ344764 IM0079A05
36	2.7	837	10	BF541441	BF541441 602069262
37	2.7	855	12	AZ189406	AZ189406 SP.1013_B
38	2.7	897	12	AZ530928	AZ530928 ENTBR71TR
39	2.7	948	12	CNS02799	AL184374 Tetraodon
40	2.7	1193	11	BC006015	BC006015 Homo sapi
41	2.6	77	9	AI345535	AI345535 tb69h12.x
42	2.6	118	10	BE464665	BE464665 hs86c05.x
43	2.6	147	10	BI274962	BI274962 UI-R-CX0-
44	2.6	148	10	BG403057	BG403057 602418871
45	2.6	165	10	BI403960	BI403960 MI-P-CPI-

ALIGNMENTS

RESULT	1
CNS04JBE	
LOCUS	
DEFINITION	CNS04JBE 879 bp DNA linear GSS 21-MAY-2000
ACCESSION	Tetraodon nigroviridis genome survey sequence T7 end of clone
VERSION	114M01 of library G from Tetraodon nigroviridis, genomic survey
KEYWORDS	sequence.
SOURCE	AL293315
ORGANISM	GSS; genome survey sequence.
REFERENCE	Tetraodon nigroviridis.
AUTHORS	Tetraodon nigroviridis
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
JOURNAL	1 (bases 1 to 879)
AUTHORS	Roest-Crollius.H., Jaillon.O., Dasilva.C., Fizes.C., Fisher.C., Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and Weissbach.J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 879)
TITLE	Roest-Crollius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C., Bernot.A., Fizes.C., Wincker.P., Brottier.P., Quetier.F., Saurin.W. and Weissbach.J.
JOURNAL	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
AUTHORS	Unpublished
TITLE	3 (bases 1 to 879)
JOURNAL	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
FEATURES	http://www.genoscope.cns.fr/Tetraodon.
source	Location/Qualifiers
	1..879
	/organism="Tetraodon nigroviridis"

```

/db_xref="taxon:99883"
/clone="114M01"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGL14AG011P1-end : 77"
BASE COUNT 207 a 220 c 223 g 226 t 3 others
ORIGIN

Query Match
Best Local Similarity 3.1%; Score 24; DB 12; Length 879;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 aaatgaataaaatgtaattttt 195
|||||
Db 851 AAAATGAATAAATGTTAATTTT 874

RESULT 2
CNS04FKX 885 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 106D12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL288474
VERSION 1 GI:8027010
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 885)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 885)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 885)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDAJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
Location/Qualifiers
1..885
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="106D12"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGL106DB06SP1-end :
PUC-Ori"
BASE COUNT 230 a 249 c 193 g 210 t 3 others
ORIGIN

Query Match
Best Local Similarity 3.1%; Score 24; DB 12; Length 885;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 aaatgaataaaatgtaattttt 195
|||||
Db* 26 AAAATGAATAAATGTTAATTTT 49

```

```

RESULT 3
AI991804/c
LOCUS AI991804
DEFINITION mRNA sequence.
ACCESSION AI991804
VERSION AI991804.1 GI:5838709
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1360 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
Location/Qualifiers
1..370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2489410"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 91 a 100 c 79 g 100 t
ORIGIN

```

```

Query Match
Best Local Similarity 2.9%; Score 23; DB 9; Length 370;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 37 GTGTAAGGCCAAAAA 15

RESULT 4
AA970105/c
LOCUS AA970105
DEFINITION OP65f11.s1 Soares,NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1581741 3' similar to TR:Q63742 Q63742 P34 PROTEIN. ; mRNA
sequence.
ACCESSION AA970105
VERSION AA970105.1 GI:3145618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 481)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 651 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 323.
Location/Qualifiers
1..481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1581741"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP.GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 110 a 131 c 116 g 124 t
ORIGIN

Query Match 2.9%; Score 23; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 33 GTCTAAAGGCAAAAAA 11

RESULT 5
AW182002/c
LOCUS
DEFINITION
xj69f03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2662493 3' similar to TR:Q63742 Q63742 P34 PROTEIN. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 178.
Location/Qualifiers
1..482
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone_lib="IMAGE:2662493"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP.GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 119 a 133 c 106 g 124 t
ORIGIN

Query Match 2.9%; Score 23; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 32 GTCTAAAGGCAAAAAA 10

RESULT 6
AI340638
LOCUS
DEFINITION
tb32c02.x1 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2056034 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 129)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 196 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2056034"
/clone_lib="NCI_CGAP_HSC2"
/tissue_type="stem cell 34+/38+"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: pAMP1; mRNA made from bone marrow, stem cells 34+/38+, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."
BASE COUNT 58 a 16 c 30 g 25 t

FEATURES

source

ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 129;
Best Local Similarity 100.0%; Pred. No. 3 4e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
Db 83 TGTAAGGCAAAAAA 104

RESULT 7

BF457830/c
LOCUS
DEFINITION BF457830 172 bp mRNA linear EST 01-DEC-2000
UI-M-BZ1-bky-e-11-0-UI.s1 NIH_BMAP_MHI2.S1 Mus musculus cDNA clone
ACCESSION BF457830
VERSION BF457830.1 GI:11523999
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 172)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
hippocampus tissue cDNA Library preparation. M.B. Soares Lab Clone
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
GENETICS. It should be noted that Bento Soares is generating a
small number of additional specialized non-redundant arrays of BMAP
cDNAs whose availability will be considered under appropriate and
limited collaborative arrangements. The following repetitive
elements were found in this cDNA sequence: 1-76,
>POLY_A#Simple_repeat
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source
1..172
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BZ1-bky-e-11-0-UI"
/clone_lib="NIH_BMAP_MHI2_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pMT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_MHI2_S1 library is a subtracted library derived
from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_MHI2_S1
TAG_TISSUE=hippocampus
TAG_SEQ=NAGTC"
52 a 24 c 23 g 73 t

BASE COUNT
ORIGIN

Query Match 2.8%; Score 22; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 tagaataaaaaatgaataaaat 185
Db 32 TAGAATAAAAAATGAATAAAAT 11

RESULT 8

BF611293
LOCUS
DEFINITION BF611293 333 bp mRNA linear EST 14-DEC-2000
dd78c05.y1 Wellcome CRC pcDNA1 egg Xenopus laevis cDNA clone
IMAGE:3430280 5', mRNA sequence.
ACCESSION BF611293
VERSION BF611293.1 GI:11780790
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 333)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Other_ESTS: dd78c05.x2
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
Gurdon, (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco.

FEATURES

source
1..333
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3430280"
/clone_lib="Wellcome CRC pcDNA1 egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pcDNA1; Site.1: NotI; Site.2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Library was
constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute)."
BASE COUNT 125 a 64 c 57 g 87 t
ORIGIN

Query Match 2.8%; Score 22; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
Db 307 TGTAAGGCAAAAAA 328

RESULT 9
AI706726/c
LOCUS

AI706726 349 bp mRNA linear EST 03-JUN-1999

```

DEFINITION  UI-R-AA1-aab-a-12-0-UI.s1 UI-R-AA1 Rattus norvegicus cDNA clone
ACCESSION   A1706726
VERSION     A1706726.1 GI:4994626
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 349)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT    Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoaresblue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..349
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-AA1-aab-a-12-0-UI"
                     /clone_lib="UI-R-AA1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7p3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AA1
                     library is a normalized library constructed from 16.5 dpc
                     rat atrium. The tag is a string of 5 nucleotides present
                     between the Not I site and the oligo-dT track. The
                     library was constructed as described by Bonaldo, Lennon
                     and Soares, Genome Research 6: 791-806, 1996. Tissue
                     provided by Jim Lin, Department of Biology, University of
                     Iowa.
BASE COUNT   99 a 71 c 60 g 119 t
ORIGIN
Query Match      2.8%; Score 22; DB 9; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaagggcaaaaaaataaaaaa 780
|||||
Db 26 TGTAAAGGCAAAAAAATAAAAAA 5

RESULT 10
LOCUS      A1927339/c
DEFINITION wn49604.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2448798 3',
            mRNA sequence.
ACCESSION  A1927339
VERSION    A1927339.1 GI:5663303
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DEFINITION  UI-R-AA1-aab-a-12-0-UI 3', mRNA sequence.
ACCESSION   A1706726.1 GI:4994626
VERSION     A1706726.1
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 349)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT    Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoaresblue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2448798"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7p3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from a
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7p3 vector. Library went through one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo."
BASE COUNT   137 a 45 c 55 g 137 t
ORIGIN
Query Match      2.8%; Score 22; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaagggcaaaaaaataaaaaa 780
|||||
Db 28 TGTAAAGGCAAAAAAATAAAAAA 7

RESULT 11
LOCUS      BG664539
DEFINITION DRABP08 Rat DRG Library Rattus norvegicus cDNA clone DRABP08 5',
            mRNA sequence.
ACCESSION  BG664539
VERSION    BG664539.1 GI:13886461
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 393)
AUTHORS    Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,
            Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and
            Zhang,X.
TITLE      Distinct gene expression profiles of rat dorsal root ganglion
            induced by peripheral nerve axotomy
JOURNAL    Unpublished (2001)
COMMENT    Contact: Zhang Xu
            Laboratory of Sensory System

```

Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn

This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR PRIMERS
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES

Location/Qualifiers
1. .393
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="DRBFB08"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
129 a 65 c 78 g 121 t

Query Match 2.8%; Score 22; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
|||||
DB 365 TGTAAGGCAAAAAA 386

RESULT 12

AA818458/c
LOCUS
DEFINITION
UT-R-A0-at-b-01-0-UI.s1 395 bp mRNA linear EST 03-JUL-1999
UT-R-A0-at-b-01-0-UI 3', similar to dbj|AB003515|AB003515 Rat mRNA
for GEF-2, complete cds, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
AA818458.1 GI:4228254
EST.
Norway rat.
Rattus norvegicus

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
On Feb 17, 1998 this sequence version replaced gi:2889197.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult Muscle library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.llnl.gov). IMAGE ID=1768015
Seq primer: M13 Forward
POLYA-No.

FEATURES

Location/Qualifiers
1. .395
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-A0-at-b-01-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; this library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
, liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
123 a 78 c 65 g 129 t

Query Match 2.8%; Score 22; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
|||||
DB 26 TGTAAGGCAAAAAA 5

RESULT 13

AA678990

LOCUS

DEFINITION

zf83e06.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone

IMAGE:383554 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 410)

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krizman,D., Kucaha,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin

,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 398.

FEATURES

Location/Qualifiers
1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:383554"
/clone_lib="Soares_pineal_gland_N3HPG"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTTTTTTTTT 3']
, double-stranded cDNA was size selected, ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of a modified pT7T3 vector

BASE COUNT 146 a 48 c 80 g 136 t
 ORIGIN (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. "

Query Match 2.8%; Score 22; DB 9; Length 410;
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaaaggcaaaaaa 780
 ||||||||||||||||||
 Db 116 TGTAAAGGCAAAAAA 137

RESULT 14
 AA818109/c
 LOCUS
 DEFINITION AA818109 448 bp mRNA linear EST 27-JAN-1999
 UI-R-A0-am-f-09-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
 UI-R-A0-am-f-09-0-UI 3', mRNA sequence.
 ACCESSION AA818109
 VERSION AA818109.1 GI:4198624
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 448)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT On Mar 9, 1998 this sequence version replaced gi:2946805.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dr track served to identify it as a clone from the normalized
 adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics

Seq primer: M13 Forward.
 FEATURES Location/Qualifiers
 1. 448

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-am-f-09-0-UI"
 /clone_lib="UI-R-A0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"

/note="Vector: pT73B-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung, brain
 , liver, kidney, heart, spleen, ovary, and muscle. The tag
 is a string of 3-5 nucleotides present between the Not I
 site and the oligo-dr track which allows identification of
 the library of origin of a clone within the mixture."

BASE COUNT 149 a 94 c 81 g 124 t
 ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 448;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaaaggcaaaaaa 780
 ||||||||||||||||||
 Db 26 TGTAAAGGCAAAAAA 5

RESULT 15
 AZ913310
 LOCUS
 DEFINITION AZ913310
 ACCESSION AZ913310
 VERSION AZ913310.1 GI:13232255
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 469)
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shvartsbeyn,A., Gebregorgis,E.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-167H1.TV

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@email.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html ;
 Plate: 167 row: H column: 1
 Seq primer: SP6
 Class: BAC ends.

FEATURES Location/Qualifiers
 1. 469

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-167H1"
 /clone_lib="RPCI-24"
 /sex="Male"

/cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 147 a 106 c 106 g 110 t
 ORIGIN

Query Match 2.8%; Score 22; DB 12; Length 469;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaaaggcaaaaaa 780
 ||||||||||||||||||
 Db 447 TGTAAAGGCAAAAAA 468

Search completed: May 14, 2002, 00:01:29
 Job time: 6684 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 13, 2002, 21:33:45 ; Search time 2160.77 seconds
(without alignments)
7554.112 Million cell updates/sec

Title: US-09-817-318-1
Perfect score: 780
Sequence: 1 aattactgtttcttcttaag.....taagcgcaaaaaaaaaa 780

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	780	100.0	780	6	AX260297
c 2	724.6	92.9	167021	9	AC079414
c 3	73.8	9.5	192126	2	AC009979
c 4	69.8	8.9	76485	3	AC024211
c 5	69.8	8.9	76485	3	AC024872
c 6	59.4	7.6	198036	2	AC109812
c 7	59	7.6	188972	2	AC006907
c 8	56.6	7.3	166011	2	AC019307
c 9	56.6	7.3	178273	2	AC005308
c 10	55.8	7.2	4325	3	PF38GAP
c 11	55.6	7.1	155711	3	VYIVD10
c 12	55.6	7.1	228309	2	AC103311
c 13	55	7.1	55339	9	AC093593
c 14	55	7.1	148983	2	AC019019
c 15	55	7.1	253305	3	PFMAL3P7
c 16	54.8	7.0	151137	2	AC090588
c 17	54.8	7.0	152881	2	AC025737
c 18	54.2	6.9	189812	9	AC009410
c 19	54	6.9	77835	2	PFMAL13P2_3
c 20	53.4	6.8	91798	9	HSKB152G3
c 21	53.2	6.8	7806	1	AF211141
c 22	53.2	6.8	199551	2	AC006281
c 23	53	6.8	47577	3	AF396436
c 24	53	6.8	186431	2	AC022281
c 25	52.8	6.8	1160	3	AF044867
c 26	52.8	6.8	256172	2	AC005139
c 27	52.8	6.7	1328	8	AY054704
c 28	52.8	6.7	160906	2	AC021055
c 29	52.8	6.7	166953	2	AC084211
c 30	52.8	6.7	192334	9	AC007319
c 31	52.8	6.7	198846	2	AC046198
c 32	52.8	6.7	80518	2	PFMAL13PA
c 33	52.8	6.7	164399	3	PFMAL3P6
c 34	52.8	6.7	86827	3	PFMAL3P5
c 35	52.8	6.7	98734	2	PFMAL1P2
c 36	52.8	6.7	318221	2	PFMAL13P3
c 37	51.8	6.6	318221	2	PFMAL13P3
c 38	51.8	6.6	321250	1	HPULM02
c 39	51.6	6.6	105485	9	HUAC002301
c 40	51.6	6.6	164498	9	AC020898
c 41	51.6	6.6	175599	2	AC009133
c 42	51.6	6.6	181705	2	AC023831
c 43	51.6	6.6	310779	2	AC005140
c 44	51.2	6.6	754	3	AAU35783
c 45	51.2	6.6	137889	9	AC073269

ALIGNMENTS

RESULT 1	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
DEFINITION	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
ACCESSION	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
VERSION	AX260297.1	GI:16509264	780 bp	DNA	linear	PAT 26-OCT-2001
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (sites)					
AUTHORS	Salceda S., Hu, P., Recipon, H. and Cafferkey, R.					
TITLE	Compositions and methods of diagnosing, monitoring, staging, imaging and treating mammary gland cancer					
JOURNAL	Patent: WO 0172780-A 1 04-OCT-2001;					
FEATURES	diadex, Inc. (US)					
Source	Location/Qualifiers					
	1..780					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
BASE COUNT	267 a	115 c	119 g	279 t		
ORIGIN						

```
Query Match      100.0%; Score 780; DB 6; Length 780;
Best Local Similarity 100.0%; Pred. No. 2,3e-123;
Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aattactgtctcttaaaagtaggagccttacaccctactaaaaatgtgatcaaaattttat 60
Db 1 AATTACTGTGCTCTTAAAGTAGAGGCTTACACCCTTACAAAATGTGATCAAAATTTTAT 60

QY 61 tatgaatagatgaagaagctgtactataaattatgagagtaagtattttattattatc 120
Db 61 TATGAATAGATGAAGAAGCTGTAGCTATAAATTATGAGAGTAAGTTTTATTATTATTATC 120

QY 121 caaatgtactatcatagacataatgacacttcactaaatcttagaataaaaaatgaat 180
Db 121 CAATGTAGTTCAATATAGCATTAATAGCAACTTCACATAATCTTTAGAAATAAAAATGAAT 180

QY 181 aaaaatgttaatttttggaggaatggttaatttttctacaaaattgtgtgacagcttt 240
Db 181 AAAATGTTAATTTTGGAGGAATGGTTAATTTTCTACAAAATTTGTGTGACAGCTTT 240

QY 241 acagaccttactcttcacaaatgtactgaacattacacatacaaaaagggtctctgttac 300
Db 241 ACAGACCTTACTCTTCACAAATTCAGCTGAACATTAACATCACAAAGAGGGTCTGTGTAC 300

QY 301 aaagaatagtcgaagaacttcagaaattttgacagtgactcttttctaaccttttaac 360
Db 301 AAAGAATAGTCGAAGAACTTCAGAAATTTTGGACAGTACTCTTTCTAACCCCTTAATC 360

QY 361 caaatatttaagtgtccatcgcttcttcttccaaactcattgttgaactagtttct 420
Db 361 CAAATATATTTAAGTGTCCATCGCTTCTTCCCTTTATCCAACTCATTTGTTAACTAGTTTCT 420

QY 421 tcgtgagttcccttgcctataattgaagcaggtctctgaaatcaaccccaaacgatttta 480
Db 421 TCGTGAGTTCCCTTGCCTATAATTGAAGCAGTCTCTGAAATCAACCCAACTGATTTTA 480

QY 481 tgaagcccatcgcttttgggaagatttgcaacttcggtcttgcaatctattacattgact 540
Db 481 TGAAGCCCATCGCTTTTGGGAAGATTGCACTTCGGCTTTGCAATCTATTTTACATTGACT 540

QY 541 gtactgtcattgtattgtctagattgtgactatcagtttaggacaaatcaaaaagatttaga 600
Db 541 GTACTGTCAATTGTATTGCTAGATTGTTGACTATCAGTTAGGACAATCAAAAAGATATTAGA 600

QY 601 taatggcaggagataaatcaagaagtactgtcaatacaaaagttattgtttatgggtatt 660
Db 601 TAATGGCAGGGATAAATCAAGAAGTTTACTGTCAATAACAAAGTTATGTTTATGGGTATT 660

QY 661 ttatagtgataaattcattactgagcaatttcatactatgtttttaaattctctggtgtg 720
Db 661 TTATAGTGATAAATTCATTACTGAGCAATTTTCATATCATGTTTTTAATTCCTCTGTTGT 720

QY 721 aatatggtgactctggagactcaaatataattattggtgtgaagggcaaaaaa 780
Db 721 AATATGTTGACTCTGGAGACTCAAAATATTAAATATTGTTGTTAAAGGCAAAAAA 780

RESULT 2
AC079414/c 167021 bp DNA linear PRI 04-JAN-2002
LOCUS Homo sapiens chromosome 16 clone RP11-358L22, complete sequence.
DEFINITION AC079414
ACCESSION AC079414
VERSION AC079414.5 GI:18057080
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167021)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 167021)
DOE Joint Genome Institute.
Direct Submission

Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 167021)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 4, 2002 this sequence version replaced gi:17976466.

Sequence Quality Assessment:

This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
Source

Location/Qualifiers
1..167021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-358L22"
BASE COUNT 49239 a 36098 c 35465 g 46219 t
ORIGIN

Query Match
Best Local Similarity
Matches 756; Conservative

92.9%; Score 724.6; DB 9; Length 167021;
98.3%; Pred. No. 1.6e-114;
0; Mismatches 4; Indels 9; Gaps 2;

QY 1 aattactgtctcttaaaagtaggagccttacaccctactaaaaatgtgatcaaaattttat 60

Db 94402 AATTACTGTGCTCTTAAAGTAGAGGCTTACACCCTTACAAAATGTGATCAAAATTTTAT 94343

QY 61 tatgaatagatgaagaagctgtactataaattatgagagtaagtattttattattatc 120

Db 94342 TATGAATAGATGAAGAAGCTGTAGCTATAAATTATGAGAGTAAGTTTTATTATTATTATC 94283

QY 121 caaatgtactatcatagacataatgacacttcactaaatcttagaataaaaaatgaat 180

Db 94282 CAAATGTAGTT-----CATATAGCAACTTCACATAATCTTAGAATAAAAAATGAAT 94231

QY 181 aaaaatgttaatttttggaggaatggttaatttttctacaaaatgtgtgacagcttt 240

Db 94230 AAAATGTTAATTTTGGAGGAATATGTTAATTTTCTACAAAATTTGTGTGACAGCTTT 94171

QY 241 acagaccttactcttcacaaatgtactgaacattacacatacaaaaagggtctctgttac 300

Db 94170 ACAGACCTTACTCTTCACAATTTGACTTTGAACATTAACATCACAAAGAGGGTCTGTGTAC 94111

QY 301 aaagaatagtcgaagaacttcagaaatttttgacagtgactcttttctaaccttttaac 360

Db 94110 AAAAGAATAGTCGAAGAACTTCATGAATTTTTCAGAGTGACTCTTTCTAACCCCTTAATC 94051

QY 361 caaatatttaagtgtccatcgcttcttcttccaaactcattgttgaactagtttct 420

Db 94050 CAAATATATTTAAGTGTCCATCGTCGTCCTTTATCCAACTCATTTGTTAACTAGTTTCT 93991

Db 50518 CTTCACTAGTGGACATTTTTCAGCAGGAGCATCAATAAGAGGCTATTTTTCAGAAATACCTGGA 50459

Qy 315 gaactcatgaatttttgacagtgactcttttctaaccctttaatccaataatatttaag 374

Db 50458 GGGCTTCATCAATTTCTCTAAGGCGCTCTCTTAATCCCTTTGATACACAAATCATTTCTCTT 50399

Qy 375 tgtccatcgtcttctcttccatccactgaattgttaactagttttcttctgtgagttcctt 434

Db 50398 CTGAAGCACCTGTCTGTCAACTGGCCTGTATGCTACCAATTCOCATCCATCAGATGCTT 50339

Qy 435 tgcctataatgaagcagctctctgaaatcaccccaactgattttatgaaagcccatgct 494

Db 50338 TGCTTGTAATTAGAGCCTTTCTCCAGCAAAATTTCCATCCACTTTCTGAATCCCTGCATC 50279

Qy 495 ttggaagatttgacttcgcttgcttgctgaactatctattacattgactgacttgct 547

Db 50278 TTTTCAAGATTGCAATTTCACTTTGCAATCAATTTGATTCATTTGTCATG 50226

RESULT 4

LOCUS AC024211 76485 bp DNA linear INV 03-NOV-2001

DEFINITION Caenorhabditis elegans cosmid Y76B12C, complete sequence.

ACCESSION AC024211

VERSION AC024211.1 GI:7105674

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 76485)

The C. elegans Sequencing Consortium.

Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium

Science 282 (5396), 2012-2018 (1998)

99069613

2 (bases 1 to 76485)

Cordes.M.

The sequence of C. elegans cosmid Y76B12C

Unpublished

3 (bases 1 to 76485)

Waterston.R.

Direct Submission

Unpublished

4 (bases 1 to 76485)

Waterston.R.

Direct Submission

Submitted (25-FEB-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

5 (bases 1 to 76485)

Waterston.R.

Direct Submission

Submitted (26-JUN-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

6 (bases 1 to 76485)

Waterston.R.

Direct Submission

Submitted (07-SEP-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

7 (bases 1 to 76485)

Waterston.R.

Direct Submission

Submitted (03-NOV-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder(P. Green and L. Hillier, ms in preparation).

FEATURES

source

1..76485

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="Y76B12C"

/complement(884..2485)

/gene="Y76B12C.4"

/complement(join(884..972,1020..1296,2093..2485))

/gene="Y76B12C.4"

/note="coded for by the following C. elegans cDNAs:

YK276c12.3, YK276c12.5"

/codon_start=1

/product="Hypothetical protein Y76B12C.4"

/protein_id="AAF36065.1"

/db_xref="GI:7105679"

/translation="MEAHRAEOPQCTKSGKVLKPGYDMTDGLSEIDDWAPKKKRAK

KVPSKRGPKLIDFEBSKEAEKEKEKEKTHEKNADKEKEAEKEEK

EKKPPEVEKEKEKEKEKTKDELEAPREKEMKKQSKDEGDLSEKEILK

IKKALKKEKKLRERLKEKIDKSTETKNSTIPSTEKEKSSMKTKKKAANATGSG

SAKKVRPPAPPPPTTKRKPKDIRAKKVCCTIL"

3822..5654

/gene="Y76B12C.3"

/join(3822..3941,3994..4089,4177..4240,5476..5654)

/gene="Y76B12C.3"

/note="contains similarity to human 15 kDa selenoprotein

(GB:AF051894); coded for by the following C. elegans

cDNAs: YK186f7.3, YK186f7.5, YK539a6.3, YK539a6.5,

YK640f8.3, YK640f8.5"

/codon_start=1

/product="Hypothetical protein Y76B12C.3"

/protein_id="AAF36064.2"

/db_xref="GI:14550393"

/translation="MWVIFLLAAVVPMPGVEVEYKIDVECKAAGFNPETLKCGLC

ERLSYHLETLTCLQCCKEKEFKHYPTALEVCNLRAPFOVAQFVKHDKAR

QRGGKVKVHVRQVQKALDKADEKTEKLSVEKWDITDILDFFNQWLE"

10101..10102

/note="S11 trans-splice site; see YK887b11.5"

10102..10415

/gene="Y76B12C.8"

/join(10102..10294,10351..10415)

/gene="Y76B12C.8"

/note="coded for by the following C. elegans cDNAs:

YK887b11.5, YK887b11.3"

/codon_start=1

/product="Hypothetical protein Y76B12C.8"

/protein_id="AAL00878.1"

/db_xref="GI:15487634"

/translation="MHFOIALIFATILLIPTEQSVPEGVSDDKETAEPKWNDDGD

EAMVYVPSLEDRSLYRLDRISGRYKSSFPYKGLGRK"

/complement(12049..14865)

/gene="Y76B12C.5"

/complement(12049..14865)

/gene="Y76B12C.5"

misc_feature

gene

CDS

gene

CDS

gene

CDS

/note="contains similarity to Pfam family PF00078 (Reverse transcriptase (RNA-dependent DNA polymerase)), score=57.5, E=2.9e-13, N=1"

/codon_start=1
/product="Hypothetical protein Y76B12C.5"
/protein_id="AAF36061.1"
/db_xref="GI:7105675"

/translation="MLPFLAASPLGDLVRAPAIWLGLVSPSPRTPIITKNSKTKRKRTD
GSTTROSSLQPPSSPLRSAGGOLRKTTRSELAKSHFHLGSMNAQSLGLPRLASGVVEY
LKNSTCVNLOSTREFSMASIDQPEQLLSPKTSKOLGLGFLFYKKCKDLEVINVS
WIGVGASLELEFRKVKIMVNIAPVCKSRSENDAKRFERLURAEITQURKSKFGPI
VLGGFNAATSCNDLAPWICGNVGNHNDFFNFNLSHRLIFQNSFRKRLA
SWTFAGKGVGTEIDFISNIDLVEDVSTFNNHLSHRLIFQNSFRKRLA
HAFKSRPSTGTERDWTLYDAIKDLRHATLGSYDSFVKTLRQGLVPLPAIKPHKF
SORTIMILORROVLSSVPDIARLKSQEARSKTIDITOKYDLSLEACRKRSSK
ILNKSIPAEILOYLEWPGDLSLTANRIKIVESHFKTLIYEAPTSPPSPRPNSNP
PNSDPVAILKCBVRLIETKATKSPGLNDVDAAMKNGDITVDSVDFALNDILHH
NKVPDMLADVKLIPKAKATKIDRFISLPLILSKSPSSILTRLTLESLYIDE
SOTKVPKGRCCADNIQSUTMLJEKNEFQLPLLLFIDYQAFDRISHSVSVSLER
AGADPMRMIQELMEGGQAEITVHOKLKNLCTGIRQGDSPALFSAALQAILTD
CONELAGVIGSVGRHILRLFEADDDVLLICSTPEEVOERLEILDRTSSNGLKINSK
TVLLKNKFCRSODILFNGSPIIPVPCGVILGRWIDISGSIDEISRIIRAGWALVGI
KEVLRIMPNERIILKQNVLPALLYASETWTNAGSTLRLKRTVTLGLDAEIRGMN
FNLERYLLAKQSRFAGHILRRDPF"
16413..21859

gene

/gene="Y76B12C.6"
join(16413..16851,18265..18363,19373..19580,20673..21468,
21659..21859)

CDS

/note="coded for by the following C. elegans cDNAs:
yk20166.5, yk459e5.5"

/codon_start=1
/product="Hypothetical protein Y76B12C.6"
/protein_id="AAF36066.1"
/db_xref="GI:7105680"

/translation="MGLVYLKSKMLERAFIIFKRILEYFIVHRLNHTGYKTFESA
OIMRIPIMYVAQOILIDFVMPVYERQAAFYRIEQLRLKILNPDAMRYMYTEFAQ
REAFYREARISITREFVELQSPLOEVLPEHHPNFPPTPEPDPRIKPTTHFF
AVFAGTPKPKTNEOMASLGLKPKTNEOMASRLDLAAIILAMNTNLPPIAGDO
PVDTLRLNLSNMVNPALTNFPOTKNOEOREAEFPDPMRAMAIPHDKDLVLELPQSD
CRNTYSPRVGHMYGPPGSSAAADLRVQAEHROPGEYQPKQSDTYIGLITTFE
ILNADATGKVLQYQVDHTTSPSTAGFMULFDLPVFAAQQOASAPPAPLQGLIDIA
TILQSGFAPGAAPPIPELDGPPRLGPTPLPLVPINPVETESSGLVWAKRGARDG
LGLSLGLSDLRDVLDAVPLPGGIRGAQQOQAVRRKRDPSHQOQSEENRRKM
SYEGARSUTCEIKPKNLFWLELLKKESRNSSEFQKNWAKFEFLRLKPKIKKNKFK
KVEKLDFRGLLR"

gene

23205..38006
/gene="Y76B12C.2"
join(23205..23327,23762..24752,26575..27035,31016..31424,
32544..32763,34005..34302,35754..35929,36655..36748,
37743..38006)

CDS

/note="contains similarity to xeroderma pigmentosum group C proteins; coded for by the following C. elegans cDNAs:
yk116e7.3, yk116e7.5"

/codon_start=1
/product="Hypothetical protein Y76B12C.2"
/protein_id="AAF36063.1"
/db_xref="GI:7105677"

/translation="METRRSSRLQOQTTSDLNAREEPQSEPPVVRKARGAKNEAPS
TAPMPKTKSTKPKQKQSPTEAELDEMGFELENEAEIIVQKSKNGRKLKVID
ENLRISAENGSSKSNLEENMEIDKPGNLAKKSKNGSRVVKSDKSNLVQSVPK
STINGSKVAIIEDDPIRAENGVKSKDEKPDFSAQNGSKLAONAPNRLSRPRSVT
TAKKVSVYSDQOLESSSSSELESSEDETEIRPKTSKIAKKEKSKISESS
SESPDDESSASESDPSIPGSEPRKRIKRSKSTLSSGGATKDLHWPCKSKASIA
RTKTPGSKNAAKPAKSKLMAQKQPKQPKNKLKDYETDRKLAKGERMLEYRHV
AHAYVARGKIEETTYDEYKLEHMKKAYFAGRSLLDAAVLDPVEFEKSNVKKSEK
NIENKTAGDSSSEDEWEHEFOPPIIDNIEVSDIDHEGGDDGDEGVVQKDWAIYL
ROEINKIREMENTHKLLECLLRLGLLELTRLVNVVRAIPRWDKTOQKLONELSKF
RELSRSSTTPGEKSLIEQKSGSKPAKQKSEKKAARKVVEERYNVEWYVQDFR
REKRWICVPLHKSVDPLSIHEHSASPSYFAIDNKOGICEVSOAMDCVQKDFR
RRTNPKVAVMTLFLPPFAANSRKKWMMQMRDLVKRLPTVMSEYKHNHLYALEK
DLLKFEALYPPPATQKPLQIRGHNVYPRSTVFTLQENNNWLKLSVKIGEKYKIV

KARDPRIIPVEDREDKFLDYGYWOTEKYRRPPLKNGKIPHNEYGNVYFMNEMCPDL
CTYLKLSGLVQISRLKGLQKQIPAVVWGAFDGGFTHPVIDGAIVLEKDAIDFINAMEKL

Query Match 8.9%; Score 69.8; DB 3; Length 76485;
Best Local Similarity 49.7%; Pred. No. 0.0028;
Matches 261; Conservative 0; Mismatches 257; Indels 7; Gaps 3;

Qy 58 tattatgaatagatgaaagcgtgtagctataaattatgagataagatttatattatttt 117
Db 35805 TATTTTCATTAACATGTACACGTTGCCATACCTGCTGTGCGGAATTTTACCTGAAAA 35746
Qy 118 atccaaatgtagtctcataatagcacttcacataatctcttcaataataaaatg 177
Db 35745 AAGGAAAT-TTGAAATTAATAATAAAATTTTATTTTGTGAAATTTTAAATTTT 35687
Qy 178 aataaaatgtaattttttggaggaatggttaatttttttcaaaaaatctgtgacagc 237
Db 35686 GTTTTGAATTAATTTTGTGAAATTAATTTTGTGAAATTTTAAATTTTGTGAAAT 35627
Qy 238 -tttacagacaccttactcttcacaaatgacttgacattcaacatcacaaagagggcctgt 296
Db 35626 TTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTTTAAATTTTAAATTTT 35567
Qy 297 ttcaaaagaatagtcagaagaactctcatgaatttttgacagtgactctttctaacaccttt 356
Db 35566 TGAATAATTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTAATTTT 35507
Qy 357 aatccaaatataatttaagtgtccatcgctctcttcttatacaactcattgttaactagt 416
Db 35506 TTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTAAT 35447
Qy 417 tctctgtgagctctcttgccataatgaagcagcttctgaaatcaccacaaactgat 476
Db 35446 TTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTTTGTGAAATTTTAAATTT 35392
Qy 477 ttatgaagccatgctgtttggaaagattgcactcgctggttgcactcttcaactattacatt 536
Db 35391 TTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTAAT 35332
Qy 537 gactgtacttcattgattgcttagatgtgactatcagtttagga 581
Db 35331 AATTTTGTGAAATTAATTTTGTGAAATTAATTTTGTGAAATTTTGTGAAATTTTGTGAAATTT 35287

RESULT 5
LOCUS AC024872 76485 bp DNA linear INV 01-MAR-2000
DEFINITION Caenorhabditis elegans clone Y76B12C, complete sequence.
ACCESSION AC024872
VERSION AC024872.1 GI:7140435
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 76485)
AUTHORS Waterston,R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 76485)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

FEATURES
source
1..76485
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y76B12C"
BASE COUNT 25107 a 13479 c 13044 g 24855 t
ORIGIN

```
Sequencing vector: M13; 46%
Chemistry: Dye-primer ET; 0% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Consensus program: Phrap; version 0.990319
Consensus quality: 189918 bases at least Q40
Consensus quality: 192873 bases at least Q30
Consensus quality: 194155 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1156: contig of 1156 bp in length
*    1157 1256: gap of unknown length
*    1257 2450: contig of 1194 bp in length
*    2451 2550: gap of unknown length
*    2551 3768: contig of 1218 bp in length
*    3769 3868: gap of unknown length
*    3869 5007: contig of 1139 bp in length
*    5008 5107: gap of unknown length
*    5108 7155: contig of 2048 bp in length
*    7156 7255: gap of unknown length
*    7256 8408: contig of 1153 bp in length
*    8409 8508: gap of unknown length
*    8509 12850: contig of 4342 bp in length
*    12851 12950: gap of unknown length
*    12951 19909: contig of 6959 bp in length
*    19910 20009: gap of unknown length
*    20010 27734: contig of 7725 bp in length
*    27735 27834: gap of unknown length
*    27835 48551: contig of 20717 bp in length
*    48552 48651: gap of unknown length
*    48652 72751: contig of 24100 bp in length
*    72752 72851: gap of unknown length
*    72852 106646: contig of 33795 bp in length
*    106647 106746: gap of unknown length
*    106747 147867: contig of 41121 bp in length
*    147868 147967: gap of unknown length
*    147968 198036: contig of 50069 bp in length.
FEATURES
Location/Qualifiers
1..198036
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="4"
   /clone="RP13-578N3"
1..11156
   /note="assembly_name:Contig19"
1257..2450
   /note="assembly_name:Contig20"
2551..3768
   /note="assembly_name:Contig24"
3869..5007
   /note="assembly_name:Contig26"
5108..7155
   /note="assembly_name:Contig38"
7256..8408
   /note="assembly_name:Contig39"
8509..12850
   /note="assembly_name:Contig42"
12951..19909
   /note="assembly_name:Contig43"
20010..27734
   /note="assembly_name:Contig44"
27835..48551
   /note="assembly_name:Contig45"
48652..72751
   /note="assembly_name:Contig46"
72852..106646
```


FEATURES

misc

misc

misc

misc

misc

mission.

misc

misc

misc

misc

misc

misc

misc

misc

misc

Query M
Best Lo

Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, J., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C.,
 Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louisghe, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M.,
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogulu, M., Okwuonu, G., Oradunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, L., Sodergren, E., Somaikhe, F., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, F., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 228309)
 Worley, K. C.
 Submitted
 Direct Submission
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 21, 2001 this sequence version replaced gi:17062975.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GKEY
 Center clone name: CH230-12416
 ----- Summary Statistics -----
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 206574 bases at least Q40
 Consensus quality: 213453 bases at least Q30
 Consensus quality: 218197 bases at least Q20
 Estimated insert size: 205273; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 54 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 21216: contig of 21216 bp in length
 * 21217 21316: gap of unknown length
 * 21317 35507: contig of 14191 bp in length
 * 35508 35607: gap of unknown length
 * 35608 45232: contig of 9625 bp in length
 * 45233 45332: gap of unknown length
 * 45333 55414: contig of 10082 bp in length
 * 55415 55514: gap of unknown length
 * 55515 65715: contig of 10201 bp in length
 * 65716 65815: gap of unknown length
 * 65816 75695: contig of 9880 bp in length
 * 75696 75795: gap of unknown length

* 75796 83809: contig of 8014 bp in length
 * 83810 91085: gap of unknown length
 * 91086 91185: contig of 17176 bp in length
 * 91186 91297: gap of unknown length
 * 91298 97297: contig of 6112 bp in length
 * 97298 102349: gap of unknown length
 * 102350 102449: gap of unknown length
 * 102450 108688: contig of 6239 bp in length
 * 108689 108788: gap of unknown length
 * 108789 114011: contig of 5223 bp in length
 * 114012 120192: gap of unknown length
 * 120193 120292: contig of 6081 bp in length
 * 120293 125370: gap of unknown length
 * 125371 125470: contig of 5078 bp in length
 * 125471 132566: gap of unknown length
 * 132567 132666: contig of 7096 bp in length
 * 132667 137559: gap of unknown length
 * 137560 137659: contig of 4893 bp in length
 * 137660 141610: gap of unknown length
 * 141611 141710: contig of 3951 bp in length
 * 141711 145979: gap of unknown length
 * 145980 146079: gap of unknown length
 * 146080 149939: contig of 3860 bp in length
 * 149940 150039: gap of unknown length
 * 150040 153905: contig of 3866 bp in length
 * 153906 154005: gap of unknown length
 * 154006 156574: contig of 2569 bp in length
 * 156575 156675: gap of unknown length
 * 156676 160464: contig of 3790 bp in length
 * 160465 160564: gap of unknown length
 * 160565 164301: contig of 3737 bp in length
 * 164302 164401: gap of unknown length
 * 164402 168556: contig of 4155 bp in length
 * 168557 168657: gap of unknown length
 * 168658 171720: contig of 3063 bp in length
 * 171721 171819: gap of unknown length
 * 171820 174507: contig of 2688 bp in length
 * 174508 174607: gap of unknown length
 * 174608 178199: contig of 3592 bp in length
 * 178200 178299: gap of unknown length
 * 178300 180055: contig of 1756 bp in length
 * 180056 180155: gap of unknown length
 * 180156 182612: contig of 2457 bp in length
 * 182613 182712: gap of unknown length
 * 182713 185727: contig of 3015 bp in length
 * 185728 185827: gap of unknown length
 * 185828 188739: contig of 2912 bp in length
 * 188740 188839: gap of unknown length
 * 188840 190935: contig of 2096 bp in length
 * 190936 191035: gap of unknown length
 * 191036 192644: contig of 1609 bp in length
 * 192645 192744: gap of unknown length
 * 192745 194719: contig of 1975 bp in length
 * 194720 194819: gap of unknown length
 * 194820 196523: contig of 1704 bp in length
 * 196524 196623: gap of unknown length
 * 196624 198412: contig of 1789 bp in length
 * 198413 198512: gap of unknown length
 * 198513 198847: contig of 1335 bp in length
 * 198848 199947: gap of unknown length
 * 199948 202688: contig of 2741 bp in length
 * 202689 202788: gap of unknown length
 * 202789 204565: contig of 1777 bp in length
 * 204566 204665: gap of unknown length
 * 204666 206363: contig of 1698 bp in length
 * 206364 206463: gap of unknown length
 * 206464 207903: contig of 1440 bp in length
 * 207904 208003: gap of unknown length
 * 208004 210550: contig of 2547 bp in length
 * 210551 212113: contig of 1463 bp in length

TITLE
 JOURNAL
 REFERENCE
 * AUTHORS
 TITLE
 JOURNAL
 COMMENT

RESULT	13
AC093593/c	
LOCUS	55339 bp DNA linear PRI 09-JAN-2002
DEFINITION	Homo sapiens BAC clone RP11-62B19 from 2, complete sequence.
ACCESSION	AC093593
VERSION	AC093593.3 GI:18093059
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 55339) Sulston,J.E. and Waterston,R. Toward a complete human genome sequence Genome Res. 8 (11), 1097-1108 (1998)
AUTHORS	99063792
JOURNAL	2 (bases 1 to 55339)
MEDLINE	Bielicki,L. and Koziolowicz,A. The sequence of Homo sapiens BAC clone RP11-62B19 Unpublished (2001)
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 148983)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavskiy, L., Boukhgaltier, B., Brown, A., Burkett, G., Castle, A.,
Choepei, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Melquins, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
ROY, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (29-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6730882.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: WBIR
Contact: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3818
Center clone name: 463.O.15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 135109 bases at least Q40
Consensus quality: 140835 bases at least Q30
Consensus quality: 143767 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 147083; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1085: contig of 1085 bp in length
* 1086 1185: gap of 100 bp
* 1186 2480: contig of 1295 bp in length
* 2481 2580: gap of 100 bp
* 2581 3989: contig of 1409 bp in length
* 3990 4089: gap of 100 bp
* 4090 4735: contig of 646 bp in length
* 4736 4835: gap of 100 bp
* 4836 6748: contig of 1913 bp in length
* 6749 6848: gap of 100 bp
* 6849 8462: contig of 1614 bp in length
* 8463 8562: gap of 100 bp
* 8563 10571: contig of 2009 bp in length
* 10572 10671: gap of 100 bp
* 10672 13999: contig of 3328 bp in length
* 14000 14099: gap of 100 bp
* 14100 17146: contig of 3047 bp in length
* 17147 17246: gap of 100 bp
* 17247 22683: contig of 5437 bp in length

* 22684 22783: gap of 100 bp
* 22784 26923: contig of 4140 bp in length
* 26924 27023: gap of 100 bp
* 27024 32598: contig of 5575 bp in length
* 32599 32698: gap of 100 bp
* 32699 40892: contig of 8194 bp in length
* 40893 40992: gap of 100 bp
* 40993 49578: contig of 8586 bp in length
* 49579 49678: gap of 100 bp
* 49679 63811: contig of 14133 bp in length
* 63812 63911: gap of 100 bp
* 63912 69287: contig of 5376 bp in length
* 69288 69387: gap of 100 bp
* 69388 82708: contig of 13321 bp in length
* 82709 82808: gap of 100 bp
* 82809 97023: contig of 14217 bp in length
* 97024 97123: gap of 100 bp
* 97124 113007: contig of 15882 bp in length
* 113008 113107: gap of 100 bp
* 113108 148983: contig of 35876 bp in length.

FEATURES
Location/Qualifiers

1..148983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-463O15"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1..1085

/note="assembly_fragment"

misc_feature

1186..2480

/note="assembly_fragment"

misc_feature

2581..3989

/note="assembly_fragment"

misc_feature

4090..4735

/note="assembly_fragment"

clone_end:T7

vector_side:left

misc_feature

4836..6748

/note="assembly_fragment"

misc_feature

6849..8462

/note="assembly_fragment"

misc_feature

8563..10571

/note="assembly_fragment"

misc_feature

10672..13999

/note="assembly_fragment"

misc_feature

14100..17146

/note="assembly_fragment"

misc_feature

17247..22683

/note="assembly_fragment"

misc_feature

22784..26923

/note="assembly_fragment"

misc_feature

27024..32598

/note="assembly_fragment"

misc_feature

32699..40892

/note="assembly_fragment"

misc_feature

40993..49578

/note="assembly_fragment"

misc_feature

49679..63811

/note="assembly_fragment"

misc_feature

63912..69287

/note="assembly_fragment"

clone_end:SP6

vector_side:left

69388..82708

/note="assembly_fragment"

misc_feature

82809..97025

/note="assembly_fragment"

misc_feature

97126..113007

/note="assembly_fragment"

misc_feature

113108..148983

/note="assembly_fragment"

BASE COUNT 41111 a 30021 c 30095 g 45854 t 1902 others
ORIGIN


```

len: 181 aa; Similarity to P. falciparum elongation factor
1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E():
1.1e-15, 53.2% identity in 141 aa overlap, revised:
shortened exon 2, Pfam: match to PF00736 EF1BD, EF-1
guanine nucleotide exchange domain Score 61.10"
/codon_start=1
/product="putative elongation factor"
/protein_id="CAB39068.2"
/db_xref="GI:8052274"
/db_xref="SPTREMBL:O97278"
/translation="MANTYDELVPLSYLYLQEGGNTSKIDQANTKKPKKKEVINKS
SLIIDIKPYGENDLDEVKLKVNITMEGLTGAKHKTPFAFLGPKLOVSCVIVDVL
VNDLELTETLGLNDELQOKKQMDDEYDEDEIGLVQSALIIISFNKL"
8235. .8240
/gene="PFC0870w, MAL3P7.3"
/notes="revised splice donor sequence for exon 2 of
PFC0870w"
8580. .8595
/notes="potential splice donor sequence may indicate
further splicing of PFC0870w"
join(11431. .20416,20578. .20690)
/gene="PFC0875w, MAL3P7.4"
join(11431. .20416,20578. .20690)
/gene="PFC0875w, MAL3P7.4"
/notes="PFC0875w (MAL3P7.4), hypothetical protein, len:
3032 aa, revised: added 3' exon, possibly spliced at 5,
end"
/codon_start=1
/product="hypothetical protein, PFC0875w"
/protein_id="CAB39024.2"
/db_xref="GI:8052275"
/db_xref="SPTREMBL:O97278"
/translation="MKDFLSYAENDNLNFIKLYNSEEDCLNKLVAIELKSDPKNVN
SLENDLIINHENDSNLLYLNHTNIINLKSEFNINDLYKDVTEFENFINKLI
NPHILVKNVATYFNKIDIKLNKNENHYTOYFNDRRKLLFFYNFVKNLVENKYSC
GLLSVENVYKQKNVFSYFLGSLSPQSEFGKGLYKNAFLYKNYENKQYGFPHL
YYPYVTESSDKKINKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEK
NGIQSVHNNNSINSDDOKKNNNNNNYYYNNPDGLITNVYKIRVGDYALLNSNEK
YFNDINTLNKONFNFNTVDDLSNFIYFNWYFSEFFILEYOFNSFILYNADILK
KNLLTPRYDELNEKREITEYMEYKDEKEYQVNNNEYINENYDEKGGKYTELK
LNDFTIPYKMSKMINAFDTEKKNIFRVVFLCVCLFIVNCFDINKERKINENFL
CCLKIKYIIYSWLLFYIILFFNYIITLVVYVYKCLMNYFILLFYFLFMS
LFTICMOFSNNSSINVIATFLPFLFSSFLRIIHGASNLIFVLLIPHSFCLIS
LDFILIKNIKIDYKOLFIFKFNISLMHLIGSIVSFLLIILANIIYIKRKMKN
MLNTRYKSKSITNDRKKSQOQTYEGNIDYMSKYPHQLKDNNDPMEGSDDE
DTCVELDTLHKQVNSOMCLRLEKIRNDRRESNGHPYDIPMPLRHNNNNNNINN
LNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NGSKSTLNIITIMISKDGEIIFFKNSYKKNKKDIYMDVKNRNGLCGRICSGY
IKFVMMNRFRNCNDNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NDQIKNLMDVEYKKISIFICPLVKRDYIILDEPFIADIKTKLKFEPDKIKNNII
FICTHDIYEANNFANDIAVIKSGOIIFPKSKNSPKLIEYKFLINIOFNGLPNEEQTN
YLSQEKIMSAIKNKGVKSPKPVANNYVDGYNVGRISHESSINDKKKNDNNDDDD
NNKNSVDHIDHLPDILVNIENFTNRRDIIKKNIIKFIKGTRENNKNCFFIENHI
YCTVKINELSLKLLCLVKNKFLNLTQKTDIYYTYIYITLNEKKLLKNIQDK
DIKLEIDDPDFFQFFNFYFNNLNNLLNNNNNNNNNNNNNNNNNNNNNNNNNNNN
IINSTNDPDYDINMYNNHKKYKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
KFIDEFSYNTYNNKNNNDENDEYDNDNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NFFTYYIKPTLLKLLKDLGSSFFYWKYVPLFLISGLLIILKCVSLFGVQVNDID
YSTISSNLKOSTINYPILYKNNMNETLGYKNYLNNDNNNNNNNNNNNNNNNNNN
TKODNAEESHINNNDNDNIKOLLENNENYNNKLHENSFNYSNTINSLLQYK
LKEKINYIDPLEDNDYDSINNTLEKANHENEISLGTFAFLITEKKADNSDNKEDY
ETLETRDNDKVEVEININLFYNTYSIHSYAYYTNLSFNLSDFONILNNKSGKNI
ILDGSTVDNIKVDKVGCDMNTLLYDKENKYNLIKIDDKNKEECNANFRLSKN
VSYNNNEEDINGNINYDNNNIIYTPKNNKEENENFLVKKRKRIDTLPNEPNIK
FHEHMRDIFYINIVFLSVIFPFCVFRFKEIENRKIFENPHVHOYIHYFOILLLE
YLYFYIILCLFLVLYFDYKEFLMSFECFLLYGFNIFSLCFSSSLYLSHYILFL
FFNFIFGIIISIVYLVILSYAYNNEILINSHLVLCIFRIFDSFALSVLNIRSLC
LNVRBRWKHIDEDIMEDMSNNNFVIFGCFKVVYNQSLDNNLNISDGIQSVCEDSN
SPNTSGDFVFLINCIIYLSIVYKLLQIIPSKERTDKNESEKTYNEEIIILNNSYN
VGVRKKEQKNRYTFELKNYLSNKEYKQPKENYEKNRPFYFIQKLFNLKRGNNY
DENNTWGIYKGEKVDNSNYSIASEININNOINSEGLKNDIYNNELNDYNDPDD
NTVDIEMQISEKSNIEKMLYKETEIDKYVKGSDDDNDDEKNFNKISOKYILKKNLI

```

misc_feature

misc_feature

gene

CDS

```

Query Match      7.1%; Score 55; DB 3; Length 253305;
Best Local Similarity 46.1%; Pred. No. 0.69;
Matches 184; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 37 actaaatgtagcaaaattttattgataagatgaagaagctgtgctataaattatga 96
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198774 AATTACAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 198775

QY 97 gactaagtattttattttattttccaaatgtagttcattagcataatagcataaacttcac 156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198714 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 198655

QY 157 taaatcattagaataaaatgataaaatgtaattttttggaggaaatggtaatttt 216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198654 TGTATATTTTAAATATATATCTCTGAATATTTTATAAATTTATAGTAAACGAAAAATAAT 198595

QY 217 tctacaaaattgtgtagacagctttacagaccttactcttcacaaattgactggaacattaa 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198594 AAGAATAACGCATGCACATATATATAGAAATATATCTATGTATATATATAGTTGTATATAATAG 198535

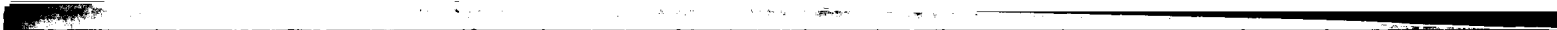
QY 277 catcacaaagagggctcgtgtttacaaaagaatagtcagaaccttcgaatttttgacag 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198534 GATAATAAATAACATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 198475

QY 337 tgactcttttctaacctttaaataataataataataataataataataataataataata 396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198474 TAATTATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 198415

QY 397 aactcattgttaactagtttcttctctgtgaggtccctt 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198414 TTTTAAATAATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 198376

```

Search completed: May 13, 2002, 22:57:25
Job time: 5020 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:35:30 ; Search time 331.61 Seconds
(without alignments)
4038.459 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgtgtctttaaag.....taaaagcaaaaaaaaaaaaaa 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	100.0	780	22	AAD19683
2	50.8	6.5	5278	22	AA546375
3	50.8	6.5	5278	24	ABL32822
4	50.2	6.4	11422	24	ABL32218
5	49.8	6.4	4590	7	AA560472
6	49.2	6.3	6179	22	AA546343
7	48.6	6.2	5516	23	ABL19458
8	48.4	6.2	17491	24	ABL34575
9	48.2	6.2	6113	24	ABL32430

10	48.2	6.2	19087	24	ABL32793	Human immune syste
11	47.8	6.1	6191	24	ABL33217	Human immune syste
12	47.8	6.1	18218	24	ABL33949	Human immune syste
13	47.4	6.1	5947	22	AA546675	Tumour suppressor
14	47.4	6.1	6106	22	AA546430	Tumour suppressor
15	47.4	6.1	6106	24	ABL33473	Human immune syste
16	47.4	6.1	16918	24	ABL33617	Human immune syste
17	47.2	6.1	4163	24	ABL34330	Human immune syste
18	47	6.0	5237	24	ABL34617	Human metastasis a
19	47	6.0	12356	22	AA546509	Tumour suppressor
20	47	6.0	13326	24	ABL33713	Human immune syste
21	46.8	6.0	13377	22	AA546476	Tumour suppressor
22	46.8	6.0	13377	24	ABL33463	Human immune syste
23	46.4	5.9	912	21	AA53727	Campylobacter jej
24	46.4	5.9	7329	22	AA546673	Tumour suppressor
25	46.4	5.9	21537	24	ABL33398	Human immune syste
26	46.2	5.9	53585	20	AA520251	Borrelia burgdorfe
27	46	5.9	6988	24	ABL34441	Human immune syste
28	46	5.9	15587	24	AA561238	Human gene regulat
29	45.8	5.9	500	21	AA593655	Cat flea head and
30	45.8	5.9	6106	22	AA546429	Tumour suppressor
31	45.8	5.9	6106	24	ABL33472	Human immune syste
32	45.8	5.9	11812	22	AA545502	Chemically pretrea
33	45.8	5.9	11812	22	AA546742	Tumour suppressor
34	45.8	5.9	11812	24	ABL34119	Human immune syste
35	45.8	5.9	15872	22	AA546519	Tumour suppressor
36	45.8	5.9	18011	24	ABL32034	Human immune syste
37	45.6	5.8	5641	24	ABL33396	Human immune syste
38	45.6	5.8	5860	24	AA561372	Human gene regulat
39	45.6	5.8	6852	24	AA561249	Human gene regulat
40	45.6	5.8	15373	24	ABL32467	Human immune syste
41	45.4	5.8	6052	24	AA561389	Human gene regulat
42	45.4	5.8	6118	24	ABL33031	Human immune syste
43	45.4	5.8	8588	22	AA545470	Chemically pretrea
44	45.4	5.8	16602	24	ABL32726	Human immune syste
45	45.2	5.8	6076	22	AA546664	Tumour suppressor

ALIGNMENTS

RESULT 1

AAD19683

ID AAD19683 standard; cDNA; 780 BP.

XX AAD19683;

XX 18-DEC-2001 (first entry)

XX Human Mammary Gland Cancer Specific Gene (MSG) mam021 cDNA.

XX Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;

XX cancer; therapy; immune response; ss.

XX Homo sapiens.

XX WO200172780-A2.

XX 04-OCT-2001.

XX 26-MAR-2001; 2001WO-US09525.

XX 27-MAR-2000; 2000US-192277P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Hu P, Recipon H, Cafferkey R;

XX WPI; 2001-616468/71.

XX New isolated polynucleotide, mammary gland cancer specific gene (MSG),
XX useful for diagnosing, monitoring, staging, imaging and treating
XX mammary gland cancer -

XX

PS Claim 1; Page 87; 99pp; English.

CC The present sequence is human mammary gland cancer specific gene (MSG)
 CC cDNA. MSG is useful for diagnosing, detecting, monitoring, staging,
 CC prognosticating, imaging and treating mammary gland cancer in a patient
 CC by determining the levels of MSG in cells, tissues or bodily fluids in a
 CC patient and comparing the determined levels of MSG with levels of MSG
 CC in cells, tissues or bodily fluids from a normal human control, where a
 CC change in determined levels of MSG in the patient versus normal control
 CC is associated with the presence of mammary gland cancer. MSG is used for
 CC identifying potential therapeutic agents for use in imaging and treating
 CC mammary gland cancer. MSG antibody conjugated to a cytotoxic agent is
 CC useful for treating mammary gland cancer in a patient. MSG vaccine is
 CC useful for inducing an immune response against a MSG protein and for
 CC treating mammary gland cancer in a patient. MSG and its protein are
 CC useful as diagnostic markers for mammary gland cancer and for diagnosis
 CC and treatment of disorders of cells, tissues and organisms.

XX Sequence 780 BP; 267 A; 115 C; 119 G; 279 T; 0 other;

Query Match 100.0%; Score 780; DB 22; Length 780;
 Best Local Similarity 100.0%; Pred. No. 1.2e-146;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aattactgttcttctaaagtgagccttacacccctactataaattgtgatcaaaattttat 60

Db 1 aattactgttcttctaaagtgagccttacacccctactataaattgtgatcaaaattttat 60

QY 61 tatgaataagatgaagcgtgtagctataaaattatgagagtaagttattttattatc 120

Db 61 tatgaataagatgaagcgtgtagctataaaattatgagagtaagttattttattatc 120

QY 121 caaattgttctcataatagcataagcaacttcactaaattcttagaataaaatgaat 180

Db 121 caaattgttctcataatagcataagcaacttcactaaattcttagaataaaatgaat 180

QY 181 aaaaatttaatttttgaggaaatggttaattttttacaaaattgtgacagcttt 240

Db 181 aaaaatttaatttttgaggaaatggttaattttttacaaaattgtgacagcttt 240

QY 241 acagaccttactctcacaattgacttgaaattacataacacacagagggctcgtttac 300

Db 241 acagaccttactctcacaattgacttgaaattacataacacacagagggctcgtttac 300

QY 301 aaaaagatagtcagaacttcataatttttgacagtgactcttttcaaccctttaac 360

Db 301 aaaaagatagtcagaacttcataatttttgacagtgactcttttcaaccctttaac 360

QY 361 caaatattatttaagtgccatgcttcttcttccactcattgttgaactagttttct 420

Db 361 caaatattatttaagtgccatgcttcttcttccactcattgttgaactagttttct 420

QY 421 tctgtgagttcttcttgcctataattgaagcaggttctctgaaatcaccacaaactgatt 480

Db 421 tctgtgagttcttcttgcctataattgaagcaggttctctgaaatcaccacaaactgatt 480

QY 481 tgaagcccatgcttttggaagatttgcaacttggcttggccttcaattattacattgact 540

Db 481 tgaagcccatgcttttggaagatttgcaacttggcttggccttcaattattacattgact 540

QY 541 gtacttgcattgtattgttagatttgaactatcagtttaggacaatacaaaagatttaga 600

Db 541 gtacttgcattgtattgttagatttgaactatcagtttaggacaatacaaaagatttaga 600

QY 601 taatggcgagggtataatcagaagttactgtcaatacaaaagttattttatgggtatt 660

Db 601 taatggcgagggtataatcagaagttactgtcaatacaaaagttattttatgggtatt 660

QY 661 ttatagtgataaattcattactgagcaatttcataatcattgttttaattctctgtgtgt 720

Db 661 ttatagtgataaattcattactgagcaatttcataatcattgttttaattctctgtgtgt 720

QY 721 aatatggtgactctggagactcaaatatttaattggtgtaaaagcacaataaaaaa 780
 Db 721 aatatggtgactctggagactcaaatatttaattggtgtaaaagcacaataaaaaa 780

RESULT 2

AA546375

ID AAS46375 standard; DNA; 5278 BP.

XX AAS46375;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #97.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 cancer

PS Claim 1; SEQ ID No 97; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5278 BP; 1568 A; 84 C; 1035 G; 2591 T; 0 other;

XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 191; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 11422 BP; 3867 A; 59 C; 1786 G; 5710 T; 0 other;

Query Match 6.4%; Score 50.2; DB 24; Length 11422;
 Best Local Similarity 55.4%; Pred. No. 0.4;
 Matches 97; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 QY 49 tcaaaattttattatgaatgatgaaagctgtagctataaaattatgagagtgatttat 108
 DB 1554 taaattcattattattattataaagtaataagagataataatgaattataatattat 1613
 QY 109 ttatatattatcaaatgttagtcaataatagcacaatagcaacttcaactaaattcttagaa 168
 DB 1614 atgataatattattatagtaataataattttaatagggaaatttaattataagataa 1673
 QY 169 taaaaatgaataaaatgttcaattttttggaggagaaatggttaattttttctacaa 223
 DB 1674 aaagtaagaagaatgatgaatgtttgtattgtattattattattattattatgtaa 1728

RESULT 5
 AAN60472/C
 ID AAN60472 standard; DNA; 4590 BP.
 XX
 AC AAN60472;
 XX
 DT 24-AUG-1991 (first entry)
 XX
 DE Sequence encoding the ring-infected Erythrocyte Surface Antigen
 DE (RESA).
 XX
 KW Malaria vaccine; antigen; epitope; ss.
 XX
 OS Plasmodium falciparum.

XX
 FH Key Location/Qualifiers
 FT exon 801..995
 FT /*tag= a
 FT 1199..4225
 FT /*tag= b
 XX
 PN WO8601802-A.
 XX
 PD 27-MAR-1986.
 XX
 PF 11-SEP-1985; 85WO-0006960.
 XX
 PR 11-SEP-1984; 84AU-0007067.
 PR 11-SEP-1984; 84AU-0007066.
 PR 10-SEP-1985; 85AU-0047326.
 XX
 PA (HALL-) HALL INST MED RES.
 XX
 PI Kemp DU, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
 XX
 XX

DR WPI; 1986-094065/14.
 DR P-PSDB; AAP60569.
 XX
 PT DNA coding for Plasmodium falciparum antigens - expressing
 PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
 PT falciparum
 XX
 PS Claim 4; Fig 1; 55pp; English.
 XX
 CC The inventors claim a novel DNA molecule which comprises a
 CC nucleotide sequence corresp. to all or a portion of the base
 CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
 CC antigenicity suitable for providing protective immunity against
 CC Plasmodium falciparum malarial infections.
 XX
 SQ Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T; 0 other;

Query Match 6.4%; Score 49.8; DB 7; Length 4590;
 Best Local Similarity 42.7%; Pred. No. 0.44;
 Matches 255; Conservative 0; Mismatches 342; Indels 0; Gaps 0;
 QY 39 taaaatgtgatcaaaattttattatgaatgatgaaagctgtagctataaaattatgaga 98
 DB 598 TAGAAAACTGTTATTTATATAAAATATATAAAAAAATAATATATAAATATT 539
 QY 99 gtaagttattttattattatccaaatgtagtctaatatagcataatagcaacttcaacta 158
 DB 538 TTAATAATAATAAATAATATTTTAAATAATATTTATTTATTTTTCGCTA 479
 QY 159 aaatctagaataaaaatgaataatgttaattttttggaggaaatggttaatttttc 218
 DB 478 TCATATAAATAAATAAATAAATAATATTTATATAAATAAATAATTTCTTTT 419
 QY 219 tacaataattgtgtgacagctttacagaccttactcttcaaatgacttgacattaca 278
 DB 418 TTTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 359
 QY 279 Tcaaaaagaggtctgttttcaaaaagaatagtcagaagacttcaatgtttgacagtg 338
 DB 358 TGAACCTTTATTTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 299
 QY 339 actcttttcaaccttttaataataataatatttaagtgtccatctcttcttcttccaa 398
 DB 298 ATAAAAAGAACAAATATATAACCAACCAAGAAATAAGGAAGCTTTTAAATTTATTATT 239
 QY 399 ctcaattgttaactagttttcttctgtgtgagtcctttgtcctataatgaagcagttctct 458
 DB 238 TTTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 179
 QY 459 gaaatcaccccaactgatttttatgaagcccatgcttttggaaagatttgcacttcgct 518
 DB 178 CGAAATCTATCACCGATATATATAATCTATATAAATAAATTTAAATAAATTTATT 119
 QY 519 ttgcaatctattacattgactgtactgtctgtattgttagattgttgcactacagtta 578
 DB 118 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 59
 QY 579 ggacaatcaaaaagatattagataatggcaggagataaaatcagaagttactgtcaat 635
 DB 58 ATAAAAAATAATATAAGTAAAAAGAAAAAGAAAAAGAAAAAGGATTCGGAAT 2

RESULT 6
 AAS46343/C
 ID AAS46343 standard; DNA; 6179 BP.
 XX
 AC AAS46343;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #65.
 XX

KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
XX
PR 06-APR-2000; 2000DE-1019058.
XX
PR 07-APR-2000; 2000DE-1019173.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
PI WPI; 2001-602752/68.
XX
DR
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX PT analysing diseases associated with cytosine methylation state e.g.
XX cancer -
XX
XX Claim 1; SEQ ID NO 65; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (ss) and sequences complementary to (Ss). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6179 BP; 1761 A; 105 C; 1302 G; 3011 T; 0 other;

Query Match 6.3%; Score 49.2; DB 22; Length 6179;
Best Local Similarity 51.4%; Pred. No. 0.59;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 5 acttgctctctaaagtaggccttacacctactaaaatgtagatcaaaattttattatg 64
DB 2832 AATTATTCCTTAATAATAATACATTCTCTAAAATCACATTATATAAAAATTTCTTTAA 2773
QY 65 aatagatgaaagcctgtagctataaaattatgagagtaagttttatttatcccaaa 124
DB 2772 AACTATTTTAACTTCCAAATTTCAATTTTCAAAATAAATCAATTCATTTCCCTTATTTT 2713
QY 125 ttagtgcataatagcatatagcaactcactcaatcttagaataaaaaatgaataaaa 184
DB 2712 TTAATCTATATTAAAAAATAACAAATTAATTTAAAAAATAAAAAAATAAAAAATAA 2653

QY 185 tgttaatttttgaggaatggtaattttttttttttttttacaaaat 226
DB 2652 AATAAACTTTAAAAAACAACACCCCACTTCATCTTCAAACT 2611
RESULT 7
ABLI9458/C
ID ABLI9458 standard; DNA: 5516 BP.
XX
XX ABLI9458;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 9847.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
KW
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 9847; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5516 BP; 1766 A; 885 C; 914 G; 1951 T; 0 other;

Query Match 6.2%; Score 48.6; DB 23; Length 5516;
Best Local Similarity 46.0%; Pred. No. 0.77;
Matches 165; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
QY 19 agtaagcgcccttacacctactaaaatgtagatcaaaattttattatgaatagatgaaagc 78
DB 365 AGTAAATTCCTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 306
QY 79 ttagctataaaattatgagagtaagtttttttttttttttttttttttttttttttt 138
DB 305 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 246
QY 139 gcaataatgcaacttcactcaatcttagaataaaaaatgaataaaatgttaattttgg 198
DB 245 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 186


```

XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1190; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 6191 BP; 1958 A; 36 C; 987 G; 3210 T; 0 other;
SQ
Query Match 6.1%; Score 47.8; DB 24; Length 6191;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 130; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 79 ttagctataaattatagagagtagttattttatttatttcccaaatgtagtccaataa 138
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4596 TATATTAAATTAACAATTAATAATACATTTTAAATATATTAATAATACATCATTA 4537
QY 139 gcaataatagcaacttcactaaacttagaataaaagaataaaatgttaattttgg 198
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4536 AAATATTACCTATTTTCTCCCTTTTAAATATATCAATTAATAAACTTAAATACATATA 4477
QY 199 aggaatgggttaattttttacaaaattgtgcagagctttacagaccttactctcac 258
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4476 TAATCTATATATACATTTTATAAAATATATATTAACCTTAAATAATACATATATC 4417
QY 259 aattgacctgaacattacaacacagaagggctcgtttacaaaagaatagccaagaac 318
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4416 CTATTTTACAACTATAAAACTTAAATATATATAAAATTAACCATACATAATAATAAC 4357
QY 319 tcatgaatttttgacagtgactcttt 345
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4356 ATCCTTAAATTTTAAACCTAACCAAT 4330
RESULT 12
ABL33949
ID ABL33949 standard; DNA; 18218 BP.
XX
XX ABL33949;
XX
XX 26-MAR-2002 (first entry)
DT
XX Human immune system associated gene SEQ ID NO: 1922.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antiataemic; cytotstatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS

```

```

XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1922; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 18218 BP; 5971 A; 315 C; 3109 G; 8823 T; 0 other;
SQ
Query Match 6.1%; Score 47.8; DB 24; Length 18218;
Best Local Similarity 46.1%; Pred. No. 1.3;
Matches 196; Conservative 0; Mismatches 227; Indels 2; Gaps 1;
QY 61 tatgaatagatgaaagctgtagctataaattatagagtagtaagttttattttatttc 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3706 ttttaaaataaattatgtttataaattatgttttaagggtattttattagatttttaa 3765
QY 121 caaatgtagtctcaataatagcataatagcaacttcactcaataatcttagaataaaaaatgaat 180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3766 atattgtattttttttgttaattagaataaaagttttataaggaatttaaaaaaat 3825
QY 181 aaaaatgttaattttttggaggaaatggttaattttttctacaaaattgtgtgacagcttt 240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3826 aaaaatttaataatagaaattaaaatgtattttttttataaaaatgtatattatta 3885
QY 241 acagaccttactcttcacaaattgacttgaacattacaacacacaaagggtctgtttac 300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3886 ataaaggatatatttaaaaaattttttgtttttataattataattatgttagaaaaatt 3945
QY 301 aaagaatagccaagaacttcacgaatttttgacagtgactcttttcaaccccttaac 360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3946 ttaatggt--tttatttttaaaatataatttttaattttttttttttttttttttttatg 4003
QY 361 caaatatttaagtgtccatgctcttccctttatccacacattttgttaactagttttct 420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4004 tttttatttaataattttgttttaaggattatttttaaggatttttttaaaatttttt 4063
QY 421 tctgtgagttcctttgctcataaattgaagcagcttctctgaatacccaaacctgatttta 480
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4064 tgtttattttttgttttttttttttttttttttttttttttttttttttttttttttt 4123
QY 481 tgaaa 485
Db ||| |||
4124 ttaaa 4128

```

RESULT 13
AAS46675

CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pt_sequences.
 XX
 SQ Sequence 6106 BP; 2270 A; 30 C; 1049 G; 2757 T; 0 other;

Query Match 6.1%; Score 47.4; DB 22; Length 6106;
 Best Local Similarity 56.5%; Pred. No. 1.4;
 Matches 108; Conservative 0; Mismatches 81; Indels 2; Gaps 1;
 QY 36 tactaaaaatgtgatcaaaattttattatgaatagatgaaagctgtagctataaaattatg 95
 Db 3205 tataaattattataaattataataataataataataa--tatatttaatttaatt 3262
 QY 96 agagtaagtgtttattttattccaaatgtgttcataatagcataaagcaacttca 155
 Db 3263 ataataatttaataataattataataataataataataataattattattattat 3322
 QY 156 cttaaatcttagaataaaaaatgaaataaattgttaatttttgggggaaatgggttaatttt 215
 Db 3323 attaatataaataataataattataataataataataataataataataataatt 3382
 QY 216 ttctacaaaaat 226
 Db 3383 taatatataat 3393

RESULT 15
 ID ABL33473 standard; DNA; 6106 BP.

XX ABL33473;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1446.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosine methylation; antiasthmatic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiarthritis; antidiabetic; antipsoriasis;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 1446; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 6106 BP; 2270 A; 30 C; 1049 G; 2757 T; 0 other;

Query Match 6.1%; Score 47.4; DB 24; Length 6106;
 Best Local Similarity 56.5%; Pred. No. 1.4;
 Matches 108; Conservative 0; Mismatches 81; Indels 2; Gaps 1;
 QY 36 tactaaaaatgtgatcaaaattttattatgaatagatgaaagctgtagctataaaattatg 95
 Db 3205 tataaattattataaattataataataataataataa--tatatttaatttaatt 3262
 QY 96 agagtaagtgtttattttattccaaatgtgttcataatagcataaagcaacttca 155
 Db 3263 ataataatttaataataattataataataataataataataattattattattat 3322
 QY 156 cttaaatcttagaataaaaaatgaaataaattgttaatttttgggggaaatgggttaatttt 215
 Db 3323 attaatataaataataataattataataataataataataataataataataatt 3382
 QY 216 ttctacaaaaat 226
 Db 3383 taatatataat 3393

Search completed: May 13, 2002, 22:53:43
 Job time: 4693 sec



Db	2063	TTATTTTGACTTTTTTTTAAATCGAGTTAAGTTGGTTATTTTCGTTATACCAATTCATAA	2122
Qy	159	aatcttagaataaaaaatgaataaaaatgttaatttttggaggaaaatggttaatttttc	218
Db	2123	AATATTTTATTATTAATTAATTAATGCATATCAGCATCTGGGTGAGTAAATATTATTA	2182
Qy	219	tacaaaatgtgtgacagctttacagacacttactcttcacaatgtgacttgaacactaaca	278
Db	2183	ACAAATATATTTTGAAAAATGATAAAAAATACTAAATGAGTTTTGGTTGAATAGTAAGA	2242
Qy	279	t	279
Db	2243	T	2243

```

RESULT      2
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgon
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	5.6%	Score 43.8;	DB 4;	Length 615;
Best Local Similarity	44.5%;	Pred. No. 0.18;		
Matches 215; Conservative	0;	Mismatches 267;	Indels 1;	Gaps 1;

Qy	1	aattactgttctctaaagtgaaggccctacacccctactaaaatgtgatgcacaaattttat	60
Db	58	ATTTAAAGTATTAAATTTATTAAACATATTATATCATTTATTATAAATTAATTATTGAT	117
Qy	61	tatgaatagatgaagaactgtagctataaattagagagtaagcttattttatatttc	120
Db	118	TATTAAATCTATTATATATAATTTATTATATAATTTACTTTAATTCATCATTTAATAATTATA	177
Qy	121	caaatgtagttcataatagcataatagcaacttcactaaactctagaataaaaaaagaat	180
Db	178	TATAAATTATAAAAAATAATTTTAATATGAATACTATTTAGTCTCTATCTCAAAATTTTAAAT	237
Qy	181	aaaatgttaatttttggagaagaatagtttaattttttctacaaaaatgtgtgacagcttt	240
Db	238	TAGTATTAAAAATATTATTAGATATTATTATTTTCTTTTAAATAAATTTAAATATAGATTAT	297
Qy	241	acgagaccttactcttcacaaattgacttgaaacttaacatcacaaagaggctcctgtttac	300
Db	298	CAATAATTAATATATTATTATTAAATGGTTATTAAAAATAATATATTTTATTATTATATA	357
Qy	301	aaaagaatagtcgaagaacttcgaatttttgacagtgactcttttctaacoccttaate	360
Db	358	GATTTAAATTTAAATATGTGAAATATT-ATTTTATTATAATATCTATTTTATAAA	416
Qy	361	caaatattttaagtgtcccatogctctctcttcacaaactcaattgttaaacatagtttct	420
Db	417	ATATTATGTGATTTATATTATTAACTTTTATAAGAAATTAATTAAATAATATTATA	476
Qy	421	tcgtgagttcccttgcctataattgaagcagdtctctcgaatacccccgaactgattta	480
Db	477	ACTTTAAATTTCTTATTATTAAATTTTTTATTATTATTTAAATAAATATATTCATTTATTAT	536
Qy	481	tga	483
Db	537	TTA	539

```

1  RESULT      3
2  US-08-299-953-1/c
3  ; Sequence 1, Application US/08299953
4  ; Patent No. 5646333
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuir
7  ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
8  ; EXPRESSION OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS
9  ; NUMBER OF SEQUENCES: 4
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
12 ; STREET: One Liberty Place 46th. Floor
13 ; CITY: Philadelphia
14 ; STATE: PA
15 ; ZIP: 19103
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Floppy disk
18 ; COMPUTER: IBM PC compatible
19 ; OPERATING SYSTEM: PC-DOS/MS-DOS
20 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: US/08/299,953
23 ; FILING DATE: Herewith
24 ; CLASSIFICATION: 435
25 ; ATTORNEY/AGENT INFORMATION:
26 ; NAME: Beardell, Lori Y.
27 ; REGISTRATION NUMBER: 34,293
28 ; REFERENCE/DOCKET NUMBER: NOVA-0003
29 ; TELECOMMUNICATION INFORMATION:
30 ; TELEPHONE: 215-564-8960
31 ; TELEFAX: 215-568-3439
32 ; INFORMATION FOR SEQ ID NO: 1:
33 ; SEQUENCE CHARACTERISTICS:
34 ; LENGTH: 2861 base pairs
35 ; TYPE: nucleic acid
36 ; STRANDEDNESS: double

```

QY	68	agatgaaagctgtagctataaaattatgagagtaagttttattttatccaaatgt	127
DB	1232	AAAAAATATCTTTATATATAATAAATGAGAAGATAAATTTATATAATTAATAAAAAA	1173
QY	128	agttcataatagcataatagcaacttcactaaatccttagaataaaaaaataaataatgt	187
DB	1172	AGATAATAATAAATTTCTAGAGTTATATAAATAAATAAATAATTAATTTTAAACATTG	1113
QY	188	taatttttggaggaaatggttaatttttttctacaaaa	225
DB	1112	TAAATGATTATATATATGATATAATTTTTTCAAAA	1075
<p>RESULT 5</p> <p>US-09-066-687-1/c</p> <p>Sequence 1, Application US/09066687</p> <p>Patent No. 6339185</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur</p> <p>TITLE OF INVENTION: A Plant Promoter Useful for Directing the</p> <p>EXPRESSION OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS</p> <p>NUMBER OF SEQUENCES: 7</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris</p> <p>STREET: One Liberty Place 46th. Floor</p> <p>CITY: Philadelphia</p> <p>STATE: PA</p> <p>ZIP: 19103</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: PatentIn Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION NUMBER: US/09/066,687</p> <p>FILING DATE: Herewith</p> <p>CLASSIFICATION:</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Beardell, Lori Y.</p> <p>REGISTRATION NUMBER: 34,293</p> <p>REFERENCE/DOCKET NUMBER: NOVA-0003</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 215-564-8960</p> <p>TELEFAX: 215-568-3439</p> <p>INFORMATION FOR SEQ ID NO: 1:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 2861 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: double</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: DNA (genomic)</p> <p>HYPOTHETICAL: NO</p> <p>ANTI-SENSE: NO</p> <p>US-09-066-687-1</p>			
<p>Query Match 5.5%; Score 42.8; DB 4; Length 2861;</p> <p>Best Local Similarity 54.4%; Pred. No. 0.38; Mismatches 0; Gaps 0;</p> <p>Matches 86; Conservative 0; Indels 72</p>			
QY	68	agatgaaagctgtagctataaaattatgagagtaagttttattttatccaaatgt	127
DB	1232	AAAAAATATCTTTATATATAATAAATGAGAAGATAAATTTATATAATTAATAAAAAA	1173
QY	128	agttcataatagcataatagcaacttcactaaatccttagaataaaaaaataaataatgt	187
DB	1172	AGATAATAATAAATTTCTAGAGTTATATAAATAAATAAATAATTAATTTTAAACATTG	1113
QY	188	taatttttggaggaaatggttaatttttttctacaaaa	225
DB	1112	TAAATGATTATATATATGATATAATTTTTTCAAAA	1075
<p>RESULT 4</p> <p>US-08-459-415-1/c</p> <p>Sequence 1, Application US/08459415</p> <p>Patent No. 5744334</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur</p> <p>TITLE OF INVENTION: A Plant Promoter Useful for Directing the</p> <p>EXPRESSION OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS</p> <p>NUMBER OF SEQUENCES: 4</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris</p> <p>STREET: One Liberty Place 46th. Floor</p> <p>CITY: Philadelphia</p> <p>STATE: PA</p> <p>ZIP: 19103</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: PatentIn Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION NUMBER: US/08/459,415</p> <p>FILING DATE: 02-JUN-1995</p> <p>CLASSIFICATION: 435</p> <p>PRIOR APPLICATION NUMBER: 08/299,953</p> <p>APPLICATION NUMBER: 08/299,953</p> <p>FILING DATE:</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Beardell, Lori Y.</p> <p>REGISTRATION NUMBER: 34,293</p> <p>REFERENCE/DOCKET NUMBER: NOVA-0003</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 215-564-8960</p> <p>TELEFAX: 215-568-3439</p> <p>INFORMATION FOR SEQ ID NO: 1:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 2861 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: double</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: DNA (genomic)</p> <p>HYPOTHETICAL: NO</p> <p>ANTI-SENSE: NO</p> <p>US-08-459-415-1</p>			
<p>Query Match 5.5%; Score 42.8; DB 1; Length 2861;</p> <p>Best Local Similarity 54.4%; Pred. No. 0.38; Mismatches 0; Gaps 0;</p> <p>Matches 86; Conservative 0; Indels 72</p>			
QY	68	agatgaaagctgtagctataaaattatgagagtaagttttattttatccaaatgt	127
DB	1232	AAAAAATATCTTTATATATAATAAATGAGAAGATAAATTTATATAATTAATAAAAAA	1173
QY	128	agttcataatagcataatagcaacttcactaaatccttagaataaaaaaataaataatgt	187
DB	1172	AGATAATAATAAATTTCTAGAGTTATATAAATAAATAAATAATTAATTTTAAACATTG	1113
QY	188	taatttttggaggaaatggttaatttttttctacaaaa	225
DB	1112	TAAATGATTATATATGATATAATTTTTTCAAAA	1075
<p>Query Match 5.5%; Score 42.8; DB 1; Length 2861;</p> <p>Best Local Similarity 54.4%; Pred. No. 0.38; Mismatches 0; Gaps 0;</p> <p>Matches 86; Conservative 0; Indels 72</p>			


```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-2

Query Match          5.5%; Score 42.8; DB 1; Length 3881;
Best Local Similarity 54.4%; Pred. No. 0.39;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagctgtagctataaaattagagagtaagtttatatttatattatccaaatgt 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1232 AAAAAAATCTTTATATATAATAAAATGGAGAGATAAATTTATATATTAATAAAAAA 1173

QY 128 agttcataatagcataatagcaactcactaaatcttagaataaaaaatgaataaatgt 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 AGATAATAATAAATCTAGAGTTATATAATAAAAACTAATATTAATCTTTTAACATTG 1113

QY 188 taatttttggaggaaatgggttaatttttctacaaaa 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 TAAAAATGATTTATATATATGATATAAATTTTTTTTCAAAA 1075

RESULT 9
US-09-066-687-2/c
; Sequence 2, Application US/09066687
; Patent No. 6339185
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,687
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-2
```

```
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-066-687-2

Query Match          5.5%; Score 42.8; DB 4; Length 3881;
Best Local Similarity 54.4%; Pred. No. 0.39;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagctgtagctataaaattagagagtaagtttatatttatattatccaaatgt 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1232 AAAAAAATCTTTATATATAATAAAATGGAGAGATAAATTTATATATTAATAAAAAA 1173

QY 128 agttcataatagcataatagcaactcactaaatcttagaataaaaaatgaataaatgt 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 AGATAATAATAAATCTAGAGTTATATAATAAAAACTAATATTAATCTTTTAACATTG 1113

QY 188 taatttttggaggaaatgggttaatttttctacaaaa 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 TAAAAATGATTTATATATATGATATAAATTTTTTTTCAAAA 1075

RESULT 10
PCT-US95-11231-2/c
; Sequence 2, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-2

Query Match          5.5%; Score 42.8; DB 5; Length 3881;
```

	Query Match	5.3%	Score 41.4;	DB 2;	Length 665;
	Best Local Similarity	45.5%	Prod. No. 0.61;		
	Matches 147;	Conservative 0;	Mismatches 176;	Indels 0;	Gaps 0;
QY	11	tcctctaaagtaagcccttacacccctactaaaatgtgatcaaaatttttatgaataga	70		
Db	172	TTTTATATTTAAATATTTTAAATTAATTAATTTTAAATTTTAAATTTTATATTTAAATA	231		
QY	71	tgaagaagctgtagctataaaattatgagagtaagttttttattttatttcccaaatgtagt	130		

```

RESULT 12
US-08-446-855A-1
; sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446.855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

```

	Query Match	Best Local Similarity	Score 41.2;	DB 2;	Length 8920;	
	Matches 88;	Conservative	0;	Mismatches	78; Indels	0; Gaps
59	attatgaatgagtagaaaagctgtgagtataaaattatgagagtaagttctttcttatatta	118				
8481	AGTATAAATATATAAATAGTAAATTTATATATAATATATATATATATATATATATA	8540				
119	tccaaatgtagtttcataatgacataatagcaacttcactaaactcttagaataaaaaatga	178				

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:32:40 ; Search time 1948.16 Seconds

(without alignments)

5403.881 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttcttcttaag.....taagggcaaaaaaaaaaaaaa 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67.2	8.6	1101	12	CNS0039G
2	62.8	8.1	1101	12	CNS00507
3	61.4	7.9	524	12	CNS01090
4	61.2	7.8	928	12	CNS000KY
c	60.8	7.8	1101	12	CNS0042W
6	60.2	7.7	1101	12	CNS00EVL
c	59.6	7.6	1101	12	CNS0039G
8	58.8	7.5	1167	12	CNS07360
c	58.2	7.5	859	12	CNS004YY
9	58	7.4	876	12	CNS009G1
11	57.2	7.3	1092	12	CNS020K7
c	57.2	7.3	1101	12	CNS00EVL
13	57.2	7.3	1101	12	CNS0022U
c	56.6	7.3	1225	12	CNS0161D
c	56.2	7.2	928	12	CNS00DKY
16	56.2	7.2	976	12	CNS04E5M
c	56.2	7.2	1101	12	CNS00507

18	55.6	7.1	1200	12	CNS016CO
c	55.2	7.1	1101	12	CNS00CYH
20	55	7.1	617	12	BH182965
21	55	7.1	617	12	CNS0708Q
c	54.6	7.0	820	12	BH1728
23	54.6	7.0	1101	12	CNS00PMC
24	54.2	6.9	475	10	BH14100
c	54.2	6.9	693	9	AV682300
26	54.2	6.9	828	12	CNS011TX
27	53.6	6.9	1084	12	CNS071NH
28	53.4	6.8	507	12	AO181894
c	53.4	6.8	987	12	CNS014FO
30	53.4	6.8	1101	12	CNS003BD
c	53.4	6.8	1190	12	CNS02ON7
32	53.2	6.8	653	12	BH527819
c	53.2	6.8	866	12	AZ668176
34	53	6.8	1101	12	CNS00238
35	53	6.8	1131	12	CNS034FO
36	53	6.8	1201	12	CNS0167M
37	52.4	6.7	914	12	CNS002JY
c	52.4	6.7	1001	12	CNS0155H
39	52.4	6.7	1001	12	CNS0155H
c	52.4	6.7	1061	12	CNS015LM
41	52.4	6.7	1086	12	CNS00YXK
c	52.4	6.7	1101	12	CNS000B8
43	52.2	6.7	859	10	BF274647
c	52.2	6.7	925	12	BH146996
c	52.2	6.7	1101	12	CNS016LI

ALIGNMENTS

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

AL063921 GI:4941778

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mamoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp; the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

FEATURES

source


```

Db 654 ATAAWAAAAAATAAAAAAATAAAAAATTTTAAWATAAAAAAATTTTATTTT 713
Qy 222 aaaaatggtgacagctttacagacaccttacttccaaatgactgacacattacatca 281
Db 714 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 773
Qy 282 caaagagggctctgtttacaaagaatagtcgaagaacttcacgaatttttgacagtgact 341
Db 774 WAAAAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 833
Qy 342 cttttcacccttttaaccataatatttaagtgctccatgctcttcttccatccaaactc 401
Db 834 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 893
Qy 402 atttgtaactagttttctctctgtgagttccttt 435
Db 894 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 927

RESULT 5
CNS0042W/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR11E08 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL055440.1 GI:4932241
VERSION AL055440
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
REFERENCE Direct Submission
AUTHORS Aaron Mammosier in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR11E08"
/note="end : T7"
BASE COUNT 294 a 74 c 99 g 381 t 253 others
ORIGIN

Query Match 7.8%; Score 60.8; DB 12; Length 1101;
Best Local Similarity 36.6%; Pred. No. 0.25;
Matches 143; Conservative 54; Mismatches 194; Indels 0; Gaps 0;

Qy 1 aattactgttctttaaagagcccttacaccctacaaatgacataaatttat 60
Db 1082 AAWAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 1023

```

```

Qy 61 tatgaatagatgaagaagctgtagctataaattatgagagtaagtttattttatttc 120
Db 1022 TAACAAAAYAWAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 963
Qy 121 caaatgagttcattatagcaacttcaactcaactcaactcaactcaactcaactcaact 180
Db 962 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 903
Qy 181 aaaaatggttaatttttggaggaaatggtaattttttctacaaaatgtgtgacagcttt 240
Db 902 TTTKATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 843
Qy 241 acagaccttactcttcacaaatgactgacattcaacataacaaagagggctctgttac 300
Db 842 TKATTTAAWATATTTTSTTATATATAATAATAATAATAATAATAATAATAATAATA 783
Qy 301 aaagaatagtcgaagaacttcacgaatttttgacagtgactcttttcaaccctttaac 360
Db 782 WATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 723
Qy 361 caaatatttgaagtgctccatgctcttcttctt 391
Db 722 TANAATTTAAWATGTTTATATATTTTATTTT 692

RESULT 6
CNS000EVL 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
REFERENCE Direct Submission
AUTHORS Aaron Mammosier in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29B23"
/note="end : T7"
BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 7.7%; Score 60.2; DB 12; Length 1101;
Best Local Similarity 34.9%; Pred. No. 0.31;

```



```

/clone="BACR19J14"
/note="end : TET3"
BASE COUNT      335 a      54 c      57 g      325 t      105 others
ORIGIN

Query Match      7.4%; Score 58; DB 12; Length 876;
Best Local Similarity 44.6%; Pred. No. 0.8;
Matches 103; Conservative 25; Mismatches 103; Indels 0; Gaps 0;

Qy 1 aattactgttctctaaagtgaagccttacacccctactaaaatggtgatcaaaatttat 60
    :||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 WATWAWTTAWATATATATATWATAAAWTTTTTAAAAATATATAAAAAATAAAWAAATAWA 541
    :||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 61 tatgaatagatgaaaagctgtagctataaaattatgagagtaagttattttatatttc 120
    : || | | : || | | : | | | | : || | | | : | | | | | : | | | |
Db 542 AAAATATATANAAAAAATTATWAWATAAATATTATTATAWAWTAAAAWTTAAWTTAAWTTATATA 601
    : || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 121 caaatgtagttcataatagcataatagcaacttcactaaatcttagaataaaaaagaat 180
    : || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 602 AWAATTTTATATATAAAWAAWATAAAAAATTATANAATTTTATATWAWWATATATATATAT 661
    : || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 181 aaaaagttaatttttggaggaaaatggttaattttctcacaaaattggt 231
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 662 WTTANWATAAATATTTTAAATATATWTTWAAATATTNNWTATATATATATATTT 712
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
CNS020K7
LOCUS
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone

```

sequence.

ACCESSION AL175696
VERSION AL175696.1 GI:7813753
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1092)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL 2 (bases 1 to 1092)
REFERENCE Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL 3 (bases 1 to 1092)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1. 1092
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"

FEATURES
source

```

/ab_xref=taxon:39003
/clone="222t11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG222CF06LP1-end : T7"
383 a 169 c 165 g 262 t 113 others

```

```

Query Match          7.3%; Score 57.2; DB 12; Length 1092;
Best Local Similarity 36.1%; Pred. No. 1;
Matches 137; Conservative 54; Mismatches 189; Indels 0; Gaps 0;

Qy 13 tctaaagtagggcctacacccctactacaaatgtgatcaaaattttattatgaatagatg 72
   ||||| : ||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 607 TTTTAAAWTTTATTTTAAWWAAAAAATAAATAAATAAATAAATAAATAAATAA 666

Qy 73 aaagctgtagctataaaatttgagagtaagttattttattattatccaaatagttc 132
   :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 667 WWWWTTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 726

Qy 133 ataagcataatagcaacttcaactacatactgaataaaatgaataaaatgttaatt 192
   ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 727 ATWAAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 786

Qy 193 ttttgaggaatggttaattttttctacaaatgtgtgacagcttttacagaccttact 252
   ||||| : ||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 787 TTTTAAWWATWAAWTTTATTTTAAWAAAAAATAAATAAATAAATAAATAAATA 846

Qy 253 cttcacattgactgacataacatacacaaagagggccctgtttacaaagaatagtc 312
   ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 847 WTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 906

Qy 313 aagaactcatgaattttgacagtgactctttctcaacccttcaatccaaatatatta 372
   :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 907 TTTTAAWWATWAAWTTTATTTTAAWAAAAAATAAATAAATAAATAAATAAATA 966

Qy 373 agtgccatcgcttccctt 392
   :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 967 WNAWAAWTTTAAATTTATTTT 986

RESULT 12
CNS00EVL          1101 bp      DNA      linear      GSS 04-JUN-1999
LOCUS             Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION        BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION         AL069706
VERSION           1
KEYWORDS          GSS.
SOURCE            fruit fly.
ORGANISM          Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE         1 (bases 1 to 1101)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                  - web : www.genoscope.cns.fr)
COMMENT           Determination of this BAC-end sequence was carried out as part of a
                  collaboration with the Berkeley Drosophila Genome Project (BDGP).
                  The BDGP is constructing a physical map of the Drosophila
                  melanogaster genome using these BACs. For further information
                  please see http://www.fruitfly.org The BDGP Drosophila
                  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                  Aaron Mammeter in Pletier de Jong's laboratory in the Department of
                  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                  NY. The library is named RPCI-98 and was constructed by partial
                  EcoRI digestion of Drosophila DNA provided by the BDGP from the
                  isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                  PI and EST libraries. A more detailed description of the library
                  and how to order individual BAC clones, the entire library, or
                  filters for hybridization from the BACPAC Resource Center can be
                  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
                  Location/Qualifiers
                  ..1101
                  /organism="Drosophila melanogaster"

/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29B23"
/notes="end : T7"

BASE COUNT      419 a      91 c      60 g      299 t      232 others
ORIGIN

Query Match          7.3%; Score 57.2; DB 12; Length 1101;
Best Local Similarity 31.7%; Pred. No. 1;
Matches 197; Conservative 105; Mismatches 319; Indels 1; Gaps 1;

Qy 41 aaatgtgatcaaaattttattatgaatagatgaaagcctgtagctataaaattatgagagt 100
   ||||| : ||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1079 ATWTTANATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1020

Qy 101 aagttattttattattatccaaatgtagttcattcaataatagcaataatagcaactcactaaa 160
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1019 TAWTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 960

Qy 161 tcttagaataaaatgaataaaatgttaattttttttggaggaatggttaatttttctta 220
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 959 T-AWATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 901

Qy 221 caaaattgtgtgacagctttacagaccccttactctccaaatgtgactgacacattcaacatc 280
   :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 900 TWWTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 841

Qy 281 acaagaggggctcctgtttacaaagaatagtcagaagaacttcacgaatttttgacagtgac 340
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 840 TATATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 781

Qy 341 tctttctcaacccttcaatccaaatatttaagtccatcgctctcccttccatccaaact 400
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 780 TWWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 721

Qy 401 cattgttaactagttcttctgtgagctccttgccttaataatgaagcagcttctctcga 460
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 720 WTATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 661

Qy 461 aatcaccacaaatgattttatgaaagcccatgcttttggaaagatttgcacctcggcttt 520
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 660 TTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 601

Qy 521 gcaatctattacattgactgactgcttgcattgtattgctagattgactatcagttagg 580
   :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 600 WAWATTTATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 541

Qy 581 acaatcaaaaagattattagataatggcaggagataaaatcagaagattactgtcacaatacaa 640
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 540 TTTTAAWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 481

Qy 641 agttatgttttatgggtatttt 662
   :||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 480 TTTWAAWATTTWKKKKKAWADT 459

RESULT 13
CNS00Z2U          1101 bp      DNA      linear      GSS 26-JUL-1999
LOCUS             Drosophila melanogaster genome survey sequence SP6 end of BAC:
DEFINITION        BACN01007 of DrosBAC library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION         AL097152
VERSION           1
KEYWORDS          GSS.
SOURCE            fruit fly.
ORGANISM          Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE         1 (bases 1 to 1101)
AUTHORS           Genoscope.

```

TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES Location/Qualifiers
 source 1..1101
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN01007"
 /note="end : SP6"
 BASE COUNT 391 a 84 c 76 g 292 t 258 others
 ORIGIN

Query Match 7.3%; Score 57.2; DB 12; Length 1101;
 Best Local Similarity 32.1%; Pred. No. 1;
 Matches 164; Conservative 76; Mismatches 269; Indels 2; Gaps 1;
 QY 270 acattacacacagaggggtcctgtttacaaaagaatagtcagaactcatgaattt 329
 Db 520 AAARKTASANNNGASASNNCTTTSSNACCACAAAKTANNNNAAAAAANNANNNCNNGC 579
 QY 330 ttgacagtgactctttcctaaaccttaaccacaaatatttaagtgtccatgctctcc 389
 Db 580 NTNNCNTNNNNANKAMAAWAATWTATMMATWTSAAATSAATSAATCTAATTVCTSA 639
 QY 390 ttatccaaactcattgttaactagttttctctgtgagttcctcttgccctataattga 449
 Db 640 TTTAAAAVATTAATT--TTAAATMATTTMTTTTATATATTTTAAAMAATTTTAAAMMTATAT 697
 QY 450 cagttctcgaatacccaactgattttatgaagcccatgctcttttgaaagatttgc 509
 Db 698 AAWTTAAAMAAMWCAAAATTTTMTATWATAAAACACAACTMTTTTAAAAAACAATNTS 757
 QY 510 acttcgcttgcaactatttaccatgactgactgactgactgactgactgactgactgact 569
 Db 758 AAAAMGSCCTTTTAATTAATVAVCVVBTVCSBSSGNNBNVYVMTTATWATAAAWT 817
 QY 570 tatcagttaggacaatcaaaaagatattagataatggcgaggataaatacagaagttact 629
 Db 818 AAWTVAATTNTRRYTAAAVAMHANNNNNWAAWMTTTTAAAAAATWAAWTTTATATT 877
 QY 630 gtcaataacaaagtattgtttatgggtattttatagggtataaattcattactgagcaa 689
 Db 878 TTDDAWMTTATATANTWMTTTTMTTWTWATAAAWMMWMAAAATTTAAWAAAAATAAA 937
 QY 690 ttccatcatgcttttaactcctcgtggttaataatgggtgactgagactcaaatatt 749
 Db 938 WTWATTWATWATTWTTTAAATTAATWAAATWAAWTTTWTWAAWNTATWATAATWT 997
 QY 750 aaatattggtgtaagggcaaaaaa 780
 Db 998 TWAAWAAAAAATATATAAAWAAAAA 1028

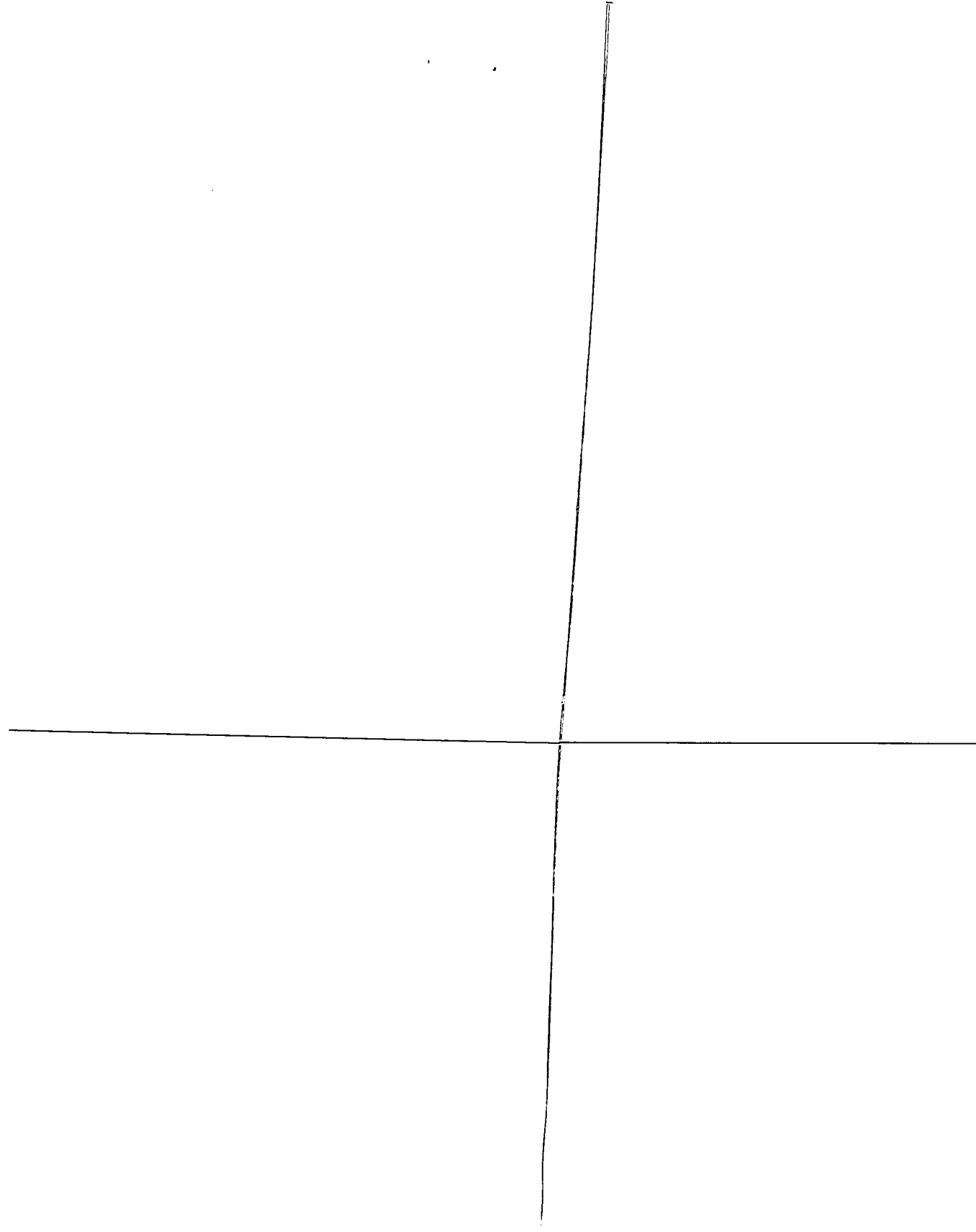
RESULT 14
CNS0161D/c 1225 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence sp6 end of BAC
DEFINITION BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106171
VERSION AL106171.1 GI:5620504

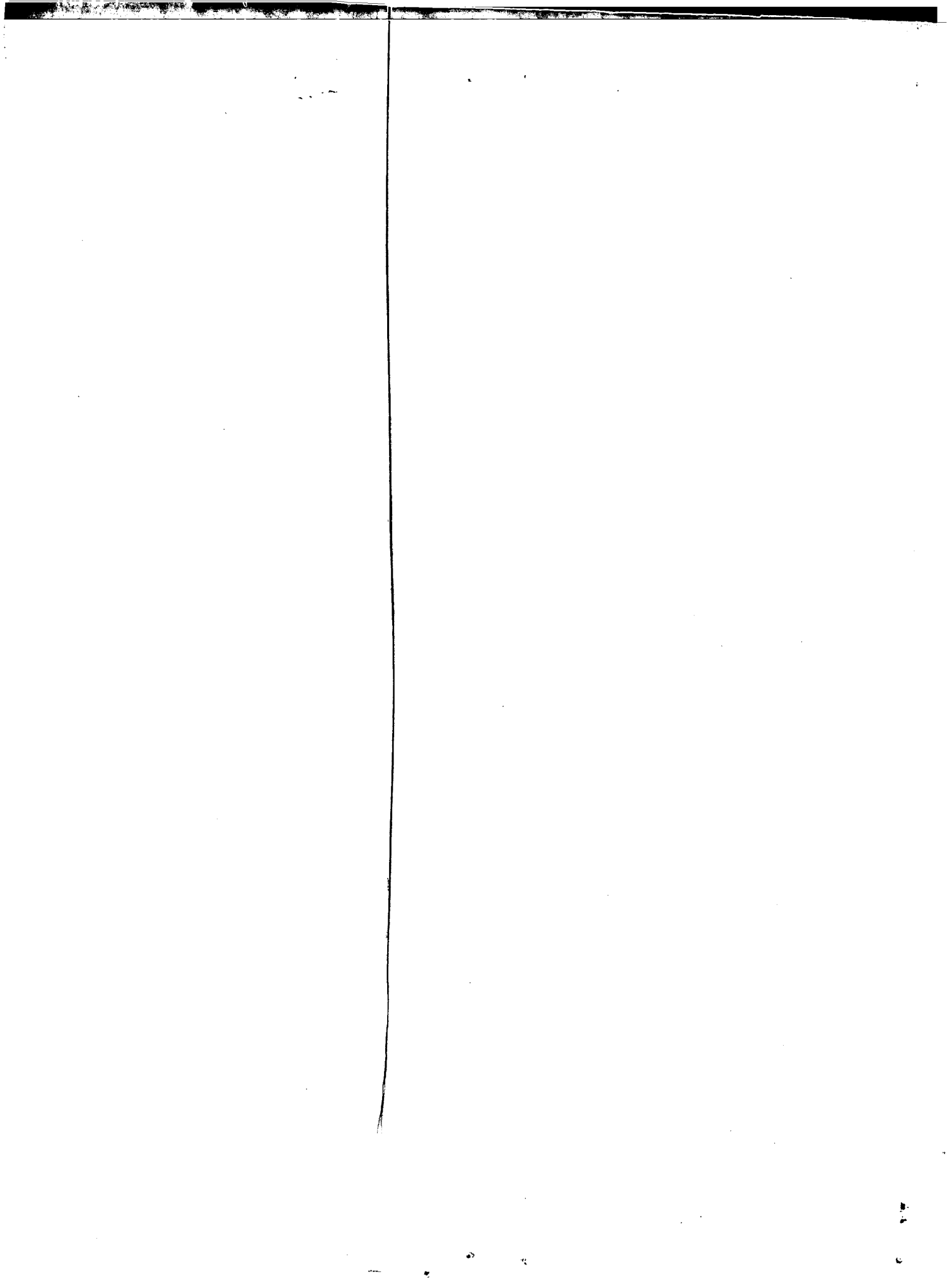
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1225)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES Location/Qualifiers
 source 1..1225
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN15C18"
 /note="end : SP6"

BASE COUNT 266 a 128 c 38 g 368 t 425 others
 ORIGIN

Query Match 7.3%; Score 56.6; DB 12; Length 1225;
 Best Local Similarity 29.2%; Pred. No. 1.3; Indels 0; Gaps 0;
 Matches 107; Conservative 90; Mismatches 170;
 QY 48 atcaaaattttattatgaatagtagaaagctgtagctataaattatgagtagtaattt 107
 Db 1103 AWWAAWAAWAAWAAWAAWATWWTWTKAATATAAAATATAATATAAAATWAAWMT 1044
 QY 108 ttttattttccaaatgtagtgcataatagcataatagcaacttcactaaattcttaga 167
 Db 1043 WAWAAAAATATATTTTTTTTWTWAAAAAATAAAATATAATATAAAWMT 984
 QY 168 ataaaaatgaataaattgtaatttttgaggagaaatggttaattttttcacaaatt 227
 Db 983 AAAAAAATAAATWTTTAAWAAWTTTTTTRAAAWMTTTTTTTTAAWMTTWTWTTTAAWAAW 924
 QY 228 gtgtgacagctttacagaccttactcttcacaaatgacttgactgaacattacacaaaga 287
 Db 923 WTTATWTRTTTTRRRRAAATWMTTTTTTTTTTTTTTTTTTTTATATTTTTTTTWHYHTY 864
 QY 288 gggctcgtttcaaaaaaatagtcagaacttcgaattttttgacagtgactctttc 347
 Db 863 TYMTTHHHMMMMHMMHMMGGGGGGTTTTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 804
 QY 348 taacctttacaaatataatlaagtgctcatcgtctctcttcttctacacactattgt 407
 Db 803 ATTTTNNRTDTWAAAAWDDTTWNNKTMTTHTTKDKMKMKGWNMMWMMGGMTRTKT 744
 QY 408 taactag 414
 Db 743 KTTKKDG 737
RESULT 15
CNS00DKY/c 928 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170





TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

Query Match 2.3%; Score 18; DB 1; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 8

US-08-832-877-49/c
Sequence 49, Application US/08832877
Patent No. 5840506
GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,877
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-877-49

Query Match 2.3%; Score 18; DB 2; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 9

US-08-832-877-49

Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1997-08-04
EARLIER APPLICATION NUMBER: US 60/054,646
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 2.3%; Score 18; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 8501 AAGGCAAAAAA 8484

RESULT 10

US-08-858-767-21
Sequence 21, Application US/08858767
Patent No. 5837468
GENERAL INFORMATION:
APPLICANT: WANG, Xun
APPLICANT: DUVICK, Jonathan P.
APPLICANT: BRIGGS, Steven P.
TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
TITLE OF INVENTION: METHOD
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,767
FILING DATE: 19-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/481,687
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/325/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136



2011



CITY: WASHINGTON
STATE: D.C.

